

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 20, 2004, 16:14:38 ; Search time 7193 Seconds
(without alignments)
10965.350 Million cell updates/sec

Title: US-09-890-811B-9
Perfect score: 1928
Sequence: 1 gacaggtctcatgcatct.....tcaaaaaaaaaaaaaaaaaa 1928

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2888711 seqs, 2045481386 residues
Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.mu.*
- 20: em.om.*
- 21: em.or.*
- 22: em.ov.*
- 23: em.pat.*
- 24: em.ph.*
- 25: em.pl.*
- 26: em.ro.*
- 27: em.sts.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htg.hum.*
- 31: em.htg.inv.*
- 32: em.htg.other.*
- 33: em.htg.mus.*
- 34: em.htg.pin.*
- 35: em.htg.rod.*
- 36: em.htg.mam.*
- 37: em.htg.vrt.*
- 38: em.sy.*
- 39: em.htgo.hum.*
- 40: em.htgo.mus.*
- 41: em.htgo.other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1928	100.0	1928	6	AX192162	AX192162 Sequence
2	549.4	28.5	1993	8	IPBSF1P	D30038 Ipomoea bat
3	489.6	25.4	78249	8	AP004529	AP004529 Lotus jap
4	454.2	23.6	1823	8	AB020590	AB020590 Nicotiana
5	430.6	22.3	1988	8	AB063573	AB063573 Nicotiana
6	424.6	22.0	1867	8	AB022693	AB022693 Nicotiana
7	388.6	20.2	1809	8	AF080595	AF080595 Pimpinell
8	379	19.7	1800	8	PCU48831	U48831 Petroselinu
9	353.4	18.3	1417	8	AB063575	AB063575 Nicotiana
10	308.2	16.0	1539	6	AX412575	AX412575 Sequence
11	308.2	16.0	1539	6	AX412905	AX412905 Sequence
12	308.2	16.0	1539	6	AX506109	AX506109 Sequence
13	308.2	16.0	1539	6	AX651760	AX651760 Sequence
14	308.2	16.0	1539	8	AF509499	AF509499 Arabidops
15	277.2	14.4	1668	6	AX654743	AX654743 Sequence
16	270.6	14.0	1213	8	BT002338	BT002338 Arabidops
17	270.6	14.0	1264	8	AY114650	AY114650 Arabidops
18	270.6	14.0	1453	8	AY062720	AY062720 Arabidops
19	270.6	14.0	1462	8	AY045813	AY045813 Arabidops
20	270.6	14.0	1606	8	AY136318	AY136318 Arabidops
21	270.2	14.0	1182	6	AX506986	AX506986 Sequence
22	270.2	14.0	1182	6	AX652016	AX652016 Sequence
23	270.2	14.0	1182	8	AF418309	AF418309 Arabidops
24	270.2	14.0	1182	8	BT008482	BT008482 Arabidops
25	237.8	12.3	1785	6	AX653781	AX653781 Sequence
26	234.6	12.2	1459	8	AFABF1	Z48429 A. fatua mRN
27	228.8	11.9	1542	6	AX505730	AX505730 Sequence
28	228.8	11.9	1542	6	AX589855	AX589855 Sequence
29	228.8	11.9	1895	8	AF224701	AF224701 Arabidops
30	221.6	11.5	2158	6	AX192164	AX192164 Sequence
31	221.4	11.5	1489	8	AY150436	AY150436 Arabidops
32	221.4	11.5	1545	6	AX507584	AX507584 Sequence
33	221.4	11.5	1545	6	AX589831	AX589831 Sequence
34	221.4	11.5	1545	8	AF425835	AF425835 Arabidops
35	221.4	11.5	1545	8	BT002629	BT002629 Arabidops
36	221.4	11.5	1736	8	AY045892	AY045892 Arabidops
37	221.4	11.5	1837	8	AF405676	AF405676 Arabidops
38	221.4	11.5	1872	8	AF425837	AF425837 Arabidops
39	221.4	11.5	1569	8	AF224703	AF224703 Arabidops
40	208.2	10.8	1626	6	AX463546	AX463546 Sequence
41	207.8	10.8	2160	8	AF096299	AF096299 Nicotiana
42	206.4	10.7	2073	8	AF140554	AF140554 Avena sat
43	204	10.6	1719	6	AX654232	AX654232 Sequence
44	187.8	9.7	1787	8	AF439274	AF439274 Retama ra
45	187.6	9.7	2086	6	AX192160	AX192160 Sequence

ALIGNMENTS

RESULT 1	AX192162	AX192162	1928 bp	DNA	linear	PAT 15-AUG-2001
LOCUS	Sequence 9 from Patent WO0149840.					
DEFINITION	Sequence 9 from Patent WO0149840.					
ACCESSION	AX192162					
VERSION	AX192162.1	GI:15210203				
KEYWORDS	Glycine max (soybean)					
SOURCE	Glycine max					
ORGANISM	Glycine max					
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;					
	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;					
	rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;					
	Glycine.					
REFERENCE	1					
AUTHORS	Zhu,Q., Famodu,O.O. and Rafalski,J.A.					

TITLE Spfl-related transcription factors		Db	
JOURNAL Patent: WO 0149840-A 9 12-JUL-2001;		Qy	
E.I. DU PONT DE NEMOURS AND COMPANY (US)		Db	
FEATURES	Location/Qualifiers	Db	
	1. .1928	Qy	
source	/organism="Glycine max"	Db	
	/mol_type="genomic DNA"	Qy	
BASE COUNT 591 a -454 c 387 g 496 t		Db	
ORIGIN		Qy	
Query Match 100.0%; Score 1928; DB 6; Length 1928;		Db	
Best Local Similarity 100.0%; Pred. No. 0;		Qy	
Matches 1928; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		Db	
Qy 1 GCACGAGTCTCATGGCACTCTCTCTGTTAGTTAGACACCTCTGCAAGTGCAGAACTCCT 60		Db 841 AAATTACTGAGATAGTTTATAAGGGTACTCATAACCCATCCTAAGCCTCAAAAATCTAGGA 900	
Db 1 GCACGAGTCTCATGGCACTCTCTCTGTTAGTTAGACACCTCTGCAAGTGCAGAACTCCT 60		Qy 901 GAAACTCATCAAACTCCTCTCTTGGCAATCCCTCATTAATTCATCAAGTGCAGAACTGAAA 960	
Qy 61 TCACCAACTTCACCTTCTCCACACACCTTTCATGACCCTCTTCTCTGACCTCTCTG 120		Db 901 GAAACTCATCAAACTCCTCTCTTGGCAATCCCTCATTAATTCATCAAGTGCAGAACTGAAA 960	
Db 61 TCACCAACTTCACCTTCTCCACACACCTTTCATGACCCTCTTCTCTGACCTCTCTG 120		Qy 961 TCCAGATCAATCCTATGCCCACACATGGAAGTGGACAAATGGATTGAGTGCACCCAG 1020	
Qy 121 CTCTCTCCCTTGGCAACAAAGCCACACAGGGTGGTTTCTCTGAGAGAACTGGCTCTG 180		Db 961 TCCAGATCAATCCTATGCCCACACATGGAAGTGGACAAATGGATTGAGTGCACCCAG 1020	
Db 121 CTCTCTCCCTTGGCAACAAAGCCACACAGGGTGGTTTCTCTGAGAGAACTGGCTCTG 180		Qy 1021 AAAACTCATCAATATCAATTTGGAGATGATTTTGGAGAGAGTTCCTCCAAAAGTGTAAAT 1080	
Qy 181 GTGTTCCCAAAATCAAGTCCACACACACCTTCTCTGCTCTCTCTCTCCCTCCCAATTT 240		Db 1021 AAAACTCATCAATATCAATTTGGAGATGATTTTGGAGAGAGTTCCTCCAAAAGTGTAAAT 1080	
Db 181 GTGTTCCCAAAATCAAGTCCACACACACCTTCTCTGCTCTCTCTCTCCCTCCCAATTT 240		Qy 1081 CAGGAGGGATGAATATGATGAAGATGAACCTGTATGCCAAAAGATGGAATAATTTGAAGTGG 1140	
Qy 241 CTCTCTCTTCTTACTTGTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 300		Db 1081 CAGGAGGGATGAATATGATGAAGATGAACCTGTATGCCAAAAGATGGAATAATTTGAAGTGG 1140	
Db 241 CTCTCTCTTCTTACTTGTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 300		Qy 1141 AAAATGAGGGTATGTGAGCCCTCTGGAAGTAGAACCTGAGAGAACCTAGAGTTGTAGTTTC 1200	
Qy 301 CGGTTCTCTTAACTCTTCCAACTCTGCTCATCTCCAACTCTGAGAGCAATTTGTTGCTC 360		Db 1141 AAAATGAGGGTATGTGAGCCCTCTGGAAGTAGAACCTGAGAGAACCTAGAGTTGTAGTTTC 1200	
Db 301 CGGTTCTCTTAACTCTTCCAACTCTGCTCATCTCCAACTCTGAGAGCAATTTGTTGCTC 360		Qy 1201 AGACAAACAGTGCATGATATCTCTTGTATGATGGCTATAGTGGAGAAAATACGGGCAGA 1260	
Qy 361 AGAGCTTCAATTTGGAAGAGAGTTCAGGGGGGAATCAGCAAAATTTGTCAAGGAAGAACA 420		Db 1201 AGACAAACAGTGCATGATATCTCTTGTATGATGGCTATAGTGGAGAAAATACGGGCAGA 1260	
Db 361 AGAGCTTCAATTTGGAAGAGAGTTCAGGGGGGAATCAGCAAAATTTGTCAAGGAAGAACA 420		Qy 1261 AAGTAGTGAAGGGCAATCCAAATCCAAAGGATTTACTACAAGTGCACACACCCAGGATGTC 1320	
Qy 421 AAGCTTCTCAAAATTTCTTCCAACTCTGCTCATCTCCAACTCTGAGAGCAATTTGTTGCTC 480		Db 1261 AAGTAGTGAAGGGCAATCCAAATCCAAAGGATTTACTACAAGTGCACACACCCAGGATGTC 1320	
Db 421 AAGCTTCTCAAAATTTCTTCCAACTCTGCTCATCTCCAACTCTGAGAGCAATTTGTTGCTC 480		Qy 1321 CAGTGAAGGAAGCAGCTGGAAAGAGCCTCACATGACCTAAGGGCTGTGATCACAACCTTATG 1380	
Qy 481 CATACCACTCTTCAAAATGTACAGTTCACCAACACAGCCATGGAGTTTTCAGGAGGCCA 540		Db 1321 CAGTGAAGGAAGCAGCTGGAAAGAGCCTCACATGACCTAAGGGCTGTGATCACAACCTTATG 1380	
Db 481 CATACCACTCTTCAAAATGTACAGTTCACCAACACAGCCATGGAGTTTTCAGGAGGCCA 540		Qy 1381 AGGAAAAGCACAACCAATGATTTCTGAGCCCTGGAGAGTGGAGAGCAATTTCTGTAACA 1440	
Qy 541 CGAAACAGGATAATTTTCTCTCAGGAAGGGTATGATGAAAATCTCTTCTTCTTCCA 600		Db 1381 AGGAAAAGCACAACCAATGATTTCTGAGCCCTGGAGAGTGGAGAGCAATTTCTGTAACA 1440	
Db 541 CGAAACAGGATAATTTTCTCTCAGGAAGGGTATGATGAAAATCTCTTCTTCTTCCA 600		Qy 1441 GACCAATGCCAAACAAATGCTTCAACACCAACCACTGAGCCACTTCGTAAGGCTCT 1500	
Qy 601 TGCAGAGTTTCCCTCTGAGTTGCTAGTGTCCAACTCAACCAAGCCATGGAGTTTTCAGGAGGCCA 660		Db 1441 GACCAATGCCAAACAAATGCTTCAACACCAACCACTGAGCCACTTCGTAAGGCTCT 1500	
Db 601 TGCAGAGTTTCCCTCTGAGTTGCTAGTGTCCAACTCAACCAAGCCATGGAGTTTTCAGGAGGCCA 660		Qy 1501 TGCCAGTGTATCCACCAAGTGCACAAATTCCTTTCAGAACCAAGATCACAAGCACCACCCAG 1560	
Qy 661 CCGATTATGGCAATTTACCCCTTCAAGTGTCCAACTCAACCAAGCCATGGAGTTTTCAGGAGGCCA 720		Db 1501 TGCCAGTGTATCCACCAAGTGCACAAATTCCTTTCAGAACCAAGATCACAAGCACCACCCAG 1560	
Db 661 CCGATTATGGCAATTTACCCCTTCAAGTGTCCAACTCAACCAAGCCATGGAGTTTTCAGGAGGCCA 720		Qy 1561 AAGGGCAATCACCTTTCAGCCCTAGAGATGCTACAAAGTCCAGGAAGTTTGGATTTCTCAG 1620	
Qy 721 ACAATTCGAGGAAATATGGCCAAACAAAGTGAAGGGAGTGAATAATTCAGAAAGTTATT 780		Db 1561 AAGGGCAATCACCTTTCAGCCCTAGAGATGCTACAAAGTCCAGGAAGTTTGGATTTCTCAG 1620	
Db 721 ACAATTCGAGGAAATATGGCCAAACAAAGTGAAGGGAGTGAATAATTCAGAAAGTTATT 780		Qy 1621 GGTTCGGATCCAATGCAATCTTACGTGAACCCAGCAGCAACTATCTGCAATGTTTCT 1680	
Qy 781 ACAATTCGACATACCCCAATTCCTTCAAGAGAGAGGTTGAGAGCTCTTTAGATGGAC 840		Db 1621 GGTTCGGATCCAATGCAATCTTACGTGAACCCAGCAGCAACTATCTGCAATGTTTCT 1680	
Db 781 ACAATTCGACATACCCCAATTCCTTCAAGAGAGAGGTTGAGAGCTCTTTAGATGGAC 840		Qy 1681 CCTCCAGGACCAAGGAGGAGCCCTAGAGATGACATGTTCTTCTTCTGAGTCTCTACTATGCTGAA 1740	
Qy 841 AAATTACTGAGATAGTTTATAAGGGTACTCATAACCCATCCTAAGCCTCAAAAATCTAGGA 900		Db 1681 CCTCCAGGACCAAGGAGGAGCCCTAGAGATGACATGTTCTTCTTCTGAGTCTCTACTATGCTGAA 1740	
Db 841 AAATTACTGAGATAGTTTATAAGGGTACTCATAACCCATCCTAAGCCTCAAAAATCTAGGA 900		Qy 1741 GGAATTTTTTTTTTCCCTTTTGGTAGCTATGAAAGTTGGAATTTTGGAGTGGGGGA 1800	
Qy 901 GAAACTCATCAAACTCCTCTCTTGGCAATCCCTCATTAATTCATCAAGTGCAGAACTGAAA 960		Db 1741 GGAATTTTTTTTTTCCCTTTTGGTAGCTATGAAAGTTGGAATTTTGGAGTGGGGGA 1800	
Db 901 GAAACTCATCAAACTCCTCTCTTGGCAATCCCTCATTAATTCATCAAGTGCAGAACTGAAA 960		Qy 1801 CTAGGATTTATTCGACAAATTAAGTTCCATTCCTATTTATTTGCAATTTTCTGTTGTTT 1860	
Qy 961 TCCAGATCAATCCTATGCCCACACATGGAAGTGGACAAATGGATTGAGTGCACCCAG 1020		Db 1801 CTAGGATTTATTCGACAAATTAAGTTCCATTCCTATTTATTTGCAATTTTCTGTTGTTT 1860	
Db 961 TCCAGATCAATCCTATGCCCACACATGGAAGTGGACAAATGGATTGAGTGCACCCAG 1020		Qy 1861 TTGTAATTTTATACAGCCACAGGATTTGGTATAGTATATATAGTATTTTCAAAAAA 1920	
Qy 1021 AAAACTCATCAATATCAATTTGGAGATGATTTTGGAGAGAGTTCCTCCAAAAGTGTAAAT 1080		Db 1861 TTGTAATTTTATACAGCCACAGGATTTGGTATAGTATATATAGTATTTTCAAAAAA 1920	
Db 1021 AAAACTCATCAATATCAATTTGGAGATGATTTTGGAGAGAGTTCCTCCAAAAGTGTAAAT 1080		Qy 1921 AAAAAAA 1928	
Qy 1081 CAGGAGGGATGAATATGATGAAGATGAACCTGTATGCCAAAAGATGGAATAATTTGAAGTGG 1140		Db 1921 AAAAAAA 1928	
Db 1081 CAGGAGGGATGAATATGATGAAGATGAACCTGTATGCCAAAAGATGGAATAATTTGAAGTGG 1140			
Qy 1141 AAAATGAGGGTATGTGAGCCCTCTGGAAGTAGAACCTGAGAGAACCTAGAGTTGTAGTTTC 1200			
Db 1141 AAAATGAGGGTATGTGAGCCCTCTGGAAGTAGAACCTGAGAGAACCTAGAGTTGTAGTTTC 1200			
Qy 1201 AGACAAACAGTGCATGATATCTCTTGTATGATGGCTATAGTGGAGAAAATACGGGCAGA 1260			
Db 1201 AGACAAACAGTGCATGATATCTCTTGTATGATGGCTATAGTGGAGAAAATACGGGCAGA 1260			
Qy 1261 AAGTAGTGAAGGGCAATCCAAATCCAAAGGATTTACTACAAGTGCACACACCCAGGATGTC 1320			
Db 1261 AAGTAGTGAAGGGCAATCCAAATCCAAAGGATTTACTACAAGTGCACACACCCAGGATGTC 1320			
Qy 1321 CAGTGAAGGAAGCAGCTGGAAAGAGCCTCACATGACCTAAGGGCTGTGATCACAACCTTATG 1380			
Db 1321 CAGTGAAGGAAGCAGCTGGAAAGAGCCTCACATGACCTAAGGGCTGTGATCACAACCTTATG 1380			
Qy 1381 AGGAAAAGCACAACCAATGATTTCTGAGCCCTGGAGAGTGGAGAGCAATTTCTGTAACA 1440			
Db 1381 AGGAAAAGCACAACCAATGATTTCTGAGCCCTGGAGAGTGGAGAGCAATTTCTGTAACA 1440			
Qy 1441 GACCAATGCCAAACAAATGCTTCAACACCAACCACTGAGCCACTTCGTAAGGCTCT 1500			
Db 1441 GACCAATGCCAAACAAATGCTTCAACACCAACCACTGAGCCACTTCGTAAGGCTCT 1500			
Qy 1501 TGCCAGTGTATCCACCAAGTGCACAAATTCCTTTCAGAACCAAGATCACAAGCACCACCCAG 1560			
Db 1501 TGCCAGTGTATCCACCAAGTGCACAAATTCCTTTCAGAACCAAGATCACAAGCACCACCCAG 1560			
Qy 1561 AAGGGCAATCACCTTTCAGCCCTAGAGATGCTACAAAGTCCAGGAAGTTTGGATTTCTCAG 1620			
Db 1561 AAGGGCAATCACCTTTCAGCCCTAGAGATGCTACAAAGTCCAGGAAGTTTGGATTTCTCAG 1620			
Qy 1621 GGTTCGGATCCAATGCAATCTTACGTGAACCCAGCAGCAACTATCTGCAATGTTTCT 1680			
Db 1621 GGTTCGGATCCAATGCAATCTTACGTGAACCCAGCAGCAACTATCTGCAATGTTTCT 1680			
Qy 1681 CCTCCAGGACCAAGGAGGAGCCCTAGAGATGACATGTTCTTCTTCTGAGTCTCTACTATGCTGAA 1740			
Db 1681 CCTCCAGGACCAAGGAGGAGCCCTAGAGATGACATGTTCTTCTTCTGAGTCTCTACTATGCTGAA 1740			
Qy 1741 GGAATTTTTTTTTTCCCTTTTGGTAGCTATGAAAGTTGGAATTTTGGAGTGGGGGA 1800			
Db 1741 GGAATTTTTTTTTTCCCTTTTGGTAGCTATGAAAGTTGGAATTTTGGAGTGGGGGA 1800			
Qy 1801 CTAGGATTTATTCGACAAATTAAGTTCCATTCCTATTTATTTGCAATTTTCTGTTGTTT 1860			
Db 1801 CTAGGATTTATTCGACAAATTAAGTTCCATTCCTATTTATTTGCAATTTTCTGTTGTTT 1860			
Qy 1861 TTGTAATTTTATACAGCCACAGGATTTGGTATAGTATATATAGTATTTTCAAAAAA 1920			
Db 1861 TTGTAATTTTATACAGCCACAGGATTTGGTATAGTATATATAGTATTTTCAAAAAA 1920			
Qy 1921 AAAAAAA 1928			
Db 1921 AAAAAAA 1928			

Qy 880 CTAAGCCTCAAACTACTAGGAGAACTCATCAAACTCTCTCTCTCTGCAATCCCTCAT 939
 Db |||||
 Qy 707 CAAGCCTCAGTCACCTAGAGA-----TCGCTTCACATCTCTTCAAAACC 754
 Db |||||
 Qy 940 CAAATTCATCAGAACTGAAATCCAGATCAATCTTATGCCACATGGAAGTGGACAA 999
 Db |||||
 Qy 755 TTGCTTACTCCAACTTGATATACAAATCAGTCAAAAGCGTCTTCTTGATAATGCTCAA 814
 Db |||||
 Qy 1000 TGAATCAGCTGCCACCCAGAAATCTCAATATCAATTTGAGATGATGATTTGAGC 1059
 Db |||||
 Qy 815 GAGATTCCTTTGCTGGACAGCAAACTCTTCGGCTCTTTTGGCGACGAGCAATGATC 874
 Db |||||
 Qy 1060 AGAGTTCCTCCAAAGTGTAAATCAGGAGGATGAATATGATGAAGATGAACCTGATGCCA 1119
 Db |||||
 Qy 875 AAGGCTCTCTATCAGTAAGTCAGGA-----GAAGATGATGAATGAACCTGAGCCAA 928
 Db |||||
 Qy 1120 AAGATGGAATTTGAAGGTGAAATCAGGATGATGTCAGCCCTCGGAAGTAGAACAGTGA 1179
 Db |||||
 Qy 929 AGAGTGAAGGTGACAAATGAAACAGAGTCATATCA---TCTGCAAGTAGAACAGTAA 985
 Db |||||
 Qy 1180 GAGAACTAGATTTGATTCAGCAACACAGTGAATGATGATGATGATGATGATGATGAT 1239
 Db |||||
 Qy 986 GAGAACTGAGATTTGATTTCAAAACCAAGCAGATGATGATTTCTTGTATGACGGTTATA 1045
 Db |||||
 Qy 1240 GGTGGAAATACGGCAGAGAACTAGTGAAGGCAATCCAAATCCAAAGGATGATGATGAT 1299
 Db |||||
 Qy 1046 GATGGAAATATGGCAAAAGTGTCCAAAGGCAATCCAAATCCAAAGGATGATGATGAT 1105
 Db |||||
 Qy 1300 AGTCACACACCCAGGATGTCAGTGAAGGACAGCTGGAAGAGCCTCAATGACCTAA 1359
 Db |||||
 Qy 1106 AATGCACATTTACTGGCTGTCAGTGAAGGACAGCTGGAAGAGCCTCAATGATCTAA 1165
 Db |||||
 Qy 1360 GGGCTGTATCAACACTTATGAGGGAAGCAACCAATGATGATGATGATGATGATGATGAT 1419
 Db |||||
 Qy 1166 GGCAGTGTATCAAACTTATGAAGGAAACACAACTGATGATGATGATGATGATGATGAT 1225
 Db |||||
 Qy 1420 GTGGCAGCCTTCTGTGAACAGCAATGCCAAATGCTTCAAAACCAACCAACCACTG 1479
 Db |||||
 Qy 1226 GTGAAGCTGTATGATGAATTAACCTTCATGAGCAACAGCAACAGATGCTGCTG 1285
 Db |||||
 Qy 1480 CAGCACTTCCGTAGGCTCTTGCCAGTATCCAAAGTGAATTCCTTCCAGTCA--- 1536
 Db |||||
 Qy 1286 TTCCGAGCCTTCAATGTGCTTAACACTCTAACCAGGGAATGAAATTTTAAACGACAT 1345
 Db |||||
 Qy 1537 ---ACCAAGATCAAGACCAACAGAGGGAATCACCCTTCACTTACCTTACAGATGCTAC 1593
 Db |||||
 Qy 1346 TTTTCAACACAGGGAACAAACACAGAACCAACCAACCACTCAGATGATGATGAT 1405
 Db |||||
 Qy 1594 AAGTCCAGGAGTTTGGATCTCAGGTTTGGGAATCCAAATGCAATCTTACGTGAAC 1653
 Db |||||
 Qy 1406 AGAGCTCTGGAATTTCAAGTTATTGAGATTTGATGATGATGATGATGATGATGATGAT 1465
 Db |||||
 Qy 1654 AGCAGCACTACTGACAAATGTTTCTCCTCCAGGACCAAGGAGGAGCTAGAGATGACA 1713
 Db |||||
 Qy 1466 AAATGCGC---CCATGATATATCAAGCGGATGAAGCAAGAGAACCTTAAAGATGAT 1522
 Db |||||
 Qy 1714 TGTTCT 1718
 Db |||||
 Qy 1523 TCTTC 1527

RESULT 7
 AF080595
 LOCUS Pimpinella brachycarpa zinc finger protein (ZFP1) mRNA, complete cds.
 DEFINITION AF080595
 ACCESSION AF080595
 VERSION AF080595.1 GI:3420905
 KEYWORDS Pimpinella brachycarpa
 SOURCE Pimpinella brachycarpa
 ORGANISM Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; campanulids; Apiales; Apiaceae; Pimpinella.
 Cho, J.-I. and Lee, K.-W.
 Pimpinella brachycarpa zinc finger protein PbZFP1 (WRKY1). mRNA
 Unpublished
 REFERENCE 2 (bases 1 to 1809)
 AUTHORS Cho, J.-I. and Lee, K.-W.
 TITLE Direct Submission
 JOURNAL Submitted (29-JUL-1998) Department of Biology, Seoul National University, Shillim-dong, Kwanak-gu, Seoul 151-742, Korea
 FEATURES
 source
 1..1809
 /organism="Pimpinella brachycarpa"
 /mol_type="mRNA"
 /db_xref="taxon:45043"
 gene
 1..1809
 /gene="ZFP1"
 195..1742
 /gene="ZFP1"
 /note="WRKY1"
 /codon_start=1
 /product="zinc finger protein"
 /protein_id="AAC31956.1"
 /db_xref="GI:3420906"
 /translation="MTSSLGLDLAQPNNDDFGSNMGFENQKTSFANPSPQFPSPSS
 PSSYFLDPIQINNYIVSSGNGTMAQNFENQFSDPFTQSRPASSTSS
 SVFANTNSVEELKRGKOGWNEFAAKNEFORPSPMTNQANMLSPMTNQANM
 QSNVAQLNSINQAQSSSQTSQKLDGYNWKYQKQVKGNSPRSYKCTY
 LNCPTKKVETFDGHTIETVYKGNHHPKQTSKSSQTSQNSIPTPTPSTLLENG
 HLEPVTTPNSLSFGEDDLFEQGNKQDDNEPDARWKRGYENNETWSSLSGR
 TVREPRIVVQTTSIDILDDGYRWKYGKQVKGNSPRSYKCTQVGCPRKHVERA
 SHDLRAVITYTEGKINHDPAPRGSGSYAVNRPSTNTATTAPTRPTNYLNPLQN
 TRAPANGQAPFTLEMLQRPSEYFSGFNGSTNTYITINQOASGQFSTAKDEPDVDS
 FFDSELA"
 BASE COUNT 629 a 399 c 326 g 455 t
 ORIGIN
 Query Match 20.2%; Score 388.6; DB 8; Length 1809;
 Best Local Similarity 65.8%; Pred. No. 2.4e-83;
 Matches 685; Conservative 0; Mismatches 299; Indels 57; Gaps 6;
 Qy 702 AGAAGGTTCAGATGATGGGTACAAATTCGAGGAAATATGGCCAAATAAAGTGAAGGAAAGT 761
 Db |||||
 Qy 756 AGTAAATTCGACGATGGATACAAATTCGAGGAAATATGGCCAAATAAAGTGAAGGAAAGT 815
 Db |||||
 Qy 762 GAAATTCGAGGAAAGTATTAATAATGCATACCCCAATTCCTTCAATGCAAGAGAGAGGTT 821
 Db |||||
 Qy 816 GAGAAATTCGAGGAAAGTATTAATAATGCATACCCCAATTCCTTCAATGCAAGAGAGAGGTT 875
 Db |||||
 Qy 822 GAGAGGTCTTTAGATGACAAATTTACTGAGATAGTTTATAAGGTTACTATAACCATCTCT 881
 Db |||||
 Qy 876 GAAACAACTTTTATGAGACACATACTGAATTTGATACAGGGGATACATCATCTCA 935
 Db |||||
 Qy 882 AAGCCTCAAAATTAAGGAGAACTCATCAAACTCTCTTCTTCTTCTTCTTCTTCTTCTTCT 941
 Db |||||
 Qy 936 AAGCCTCAAAATTAAGGAGAACTCATCTTCAATCTTATCAGAAATTCATACCTCTCT 989
 Db |||||
 Qy 942 AATTCATCAGAACTGAAATCCAGATCAATCTTATGCGCACATGGAAGTGAAGAGATG 1001
 Db |||||
 Qy 990 -----ACCATGCCAGAACTTCTTACTAGAAAATGGCCATTATA 1028
 Db |||||
 Qy 1002 GATTCAGTCCACCCAGAGAACTCATCAATATCAATTCGAGATGATGATTTTGAGCAG 1061
 Db |||||
 Qy 1029 GAACCAAGTAAACCCCGGAAACTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1088
 Db |||||
 Qy 1062 AGTTCCCAAAGTGTAAATCAGGAGGAGTGAATATGATGAAGATGAACCTGATGCGCAA 1121
 Db |||||
 Qy 1089 CAAGGATCCATGAATAAACAAGG-----GACGATGATGAATAAAGCTGATGCGCAA 1142
 Db |||||
 Qy 1122 AGATGGAATTTGAAGTG---AAATGAGGATGATGTCAGCCCTGGAAGTGAAGAGATG 1178
 Db |||||
 Qy 1143 AGATGGAAGGATGATGATGAATAATGAGCAATGTCATCTCTTGGGAGTTCGAAGCTGC 1202
 Db |||||


```
QY 1179 AGAGAACCTAGAGTTGTTAGTTCAGACAAACAGTGCATTTGATATCTTCTGATGATGCTAT 1238
Db 1203 CGCGAACCAAGATTGTTAGTTCACAAACAACTAGTATATAGATATCTTCTGATGATGCTAT 1262
QY 1239 AGGTGAGAAATATACCGGAGAAAGTAGTGAAGGCAATCCAAATCCAAAGGATTTACTATC 1298
Db 1263 AGGTGAGAAATATGAGACAGAGGTTGTTAAAGGCAATCCAAATCCAAAGGATTTACTATC 1322
QY 1299 AGTGCACACACCCAGGATGTCAGTGAGGAGACGCTGGAAGAGCCTCACATGACCTA 1358
Db 1323 AATGTACTCAAGTTGGCTGTCGGTGAGAAACACAGTGAGAGACATCTCATGATTTG 1382
QY 1359 AGGGCTGTGATCAACACTTATGAGGGAAGCAACCAATCATGATGTTCTCTGAGCCCGCTGGC 1418
Db 1383 AGGGCAGTATCACCACTTACGAAGGAACACATCAACATGATGTTCTCTGACACCGCGC 1442
QY 1419 AGTGGAGCCATCTCTGTGAACAGACCAATGCCAAACAAATGCTTCAACACACCAACACT 1478
Db 1443 AGTGGCAGTATGACAGTAAATAGACCATCTGTATAACACCGC-----CACCAACAGT 1493
QY 1479 GCAGCCACTTCCGTAGGCTCTTCCAGTGCATCCCAAGTACCAATTCCTTTCAGAAC 1538
Db 1494 GTCCACACGCTATAAGC-----CAACCAACAAATTAATTAACCCCTCTTTCAGAAC 1544
QY 1539 CAAAGATACAAAGCACCACAGAAAGGCAATCAACCTTCAACCTAGAGATGCTTACAAAGT 1598
Db 1545 ACTAGAGCACA---ACTGCAATGCGCAAGCACCATTACACTAGAAATGCTGACGCT 1601
QY 1599 CCAGGAAGTTTGGATCTCAGGTTTGGGAATCCAAATGCAATCTTACGTGAACACGACAG 1658
Db 1602 CCAGGAAGTACGAGTTCTCAGGTTTGGAACTCACTAAATTAATTAACGATCAACCAA 1661
QY 1659 CAATCTGACAAATGTTTCTCTCCAGGACCAAGGAGGAGGCTAGAGATGACATGTTTC 1718
Db 1662 AATCAACAGCAAGTGACAGTTCTCCAGGCAAGGAGGAGGCTAGAGATGACATGTTTC 1721
QY 1719 CTTGAGTCTCTACTGCTGA 1739
Db 1722 TTTGACTCTTTTGGCCTGA 1742

RESULT 8
PCU48831 1800 bp mRNA linear PLN 19-NOV-1996
LOCUS Petroselinum crispum DNA-binding protein WRKY1 mRNA, complete cds.
DEFINITION U48831
ACCESSION U48831.1 GI:1431871
VERSION 1
KEYWORDS Petroselinum crispum (parsley)
SOURCE Petroselinum crispum
ORGANISM Petroselinum crispum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
Asteridae; campanulids; Apiales; Apiaceae; Petroselinum.
REFERENCE 1 (bases 1 to 1800)
AUTHORS Rushon,P.J., Torres,J.T., Parniske,M., Wernert,P., Hahlbrock,K.
and Somasich,I.E.
TITLE Interaction of elicitor-induced DNA-binding proteins with elicitor
response elements in the promoters of parsley PR1 genes
JOURNAL EMBO J. 15 (20), 5690-5700 (1996)
MEDLINE 97051827
PUBMED 8896462
REFERENCE 2 (bases 1 to 1800)
AUTHORS Rushon,P.
TITLE Direct Submission
SUBMITTED (09-FEB-1996) Paul Rushton, Dept of Biochemistry,
Max-planck-Institut fuer Zuechtungsforschung, Carl-von-linne Weg 10,
Koeln 50829, Germany
FEATURES
Location/Qualifiers
1..1800
/organism="Petroselinum crispum"
/mol_type="mRNA"
/db_xref="taxon:4043"

52..1596
/note="contains two WRKY domains; WRKY-type DNA-binding
protein; sequence-specific DNA-binding protein"
/codon_start=1
/product="WRKY1"
/protein_id="AAC49527.1"
/db_xref="GI:1431872"
/translation="MSSSLGLDILAQNNDFFGSMNGFNQKTSFANQSLPFPSPVPS
PSYFSLDPSLQNNNTYIVSSNGKTNNAQSFKEENQNFSDPFAOTRPSASTSSS
LIPTANLVEVSLKRGKGNFEBPAKNDPLMDNASVTSDIATLQRISEPMNQAAN
MQSNAALQSNLNNYAQSQSQTNRDQSKLDGYNWKYQKQVKGSENPRSYVKCTY
LMCTPKYVETTFDGHTEIVYKGNHNPKEPOSTKRSSQSYQNSIGTNPESSLLENG
RSEPTVTPESSLSFEGDDLFQGGNKPQDDGNEPDSKRWKGYTESNEPWSLLSR
TVREPRIVVQTTSDIILDGYNWKYQKQVKGNDPRSYVKCTQVQCPVKRKHVERA
SHDLRAVTTTVEGKHNDVPAPRGSGSYPAVNRPSDNTSAPTARPTTNYLNPQNP
RAOPANGAOPFTLEMLQRPSTYFSGFTNTSNTYAINQQAQSGQSFSTAKDPEVDSE
FDSFLA"
misc_feature 340..357
/note="encodes serine/threonine-rich region"
misc_feature 493..612
/note="encodes glutamine-rich region"
misc_feature 627..789
/note="encodes WRKY domain and a putative DNA-binding
domain and zinc finger domain"
misc_feature 970..993
/note="encodes acidic domain"
misc_feature 997..1008
/note="encodes basic domain"
misc_feature 1093..1113
/note="encodes acidic domain"
misc_feature 1111..1278
/note="encodes WRKY domain"
BASE COUNT 610 a 379 c 334 g 477 t
ORIGIN
Query Match 19.7%; Score 379; DB 8; Length 1800;
Best Local Similarity 65.9%; Pred. No. 5.1e-81;
Matches 686; Conservative 0; Mismatches 295; Indels 60; Gaps 7;
QY 702 AGAAGTCAGATGATGGGTACAAATTCGAGGAATATGGCCAAAACAAAGTGAAGGAAGT 761
Db 613 AGTAAATGGAGATGGATACAAATTCGAGGAATATGGAGAAATATGGAGCAAGCAGGTTAAGGAAGT 672
QY 762 GAAATTCAGAAAGTTATTACAAATGCACATACCCCAATTCGCCCTACAAAGAGAAGAGTT 821
Db 673 GAGATTCAGAAAGTTATTACAAATGCACATATCTTAATGTCCTACTAAGAAAAGGTG 732
QY 822 GAGAGTCTTTAGATGGACAAATTAATCTGAGATAGTTTATAAGGGTACTCATACCAATCCT 881
Db 733 GAAACAACTTTTATGGACACATACTGAAATTTGTATACAAAGGCAATCATACCAATCCA 792
QY 882 AGCCCTCAAAATTAATAGGAAACTCATCAAACTCTCTCTCTCTCTCTCTCTCTCTCTCTCA 941
Db 793 AAGCCCTCAATCAACGAAACGA-----TCACTTTCACAGTCTTACCAAGAATTC 840
QY 942 AATTCATCAGAACTGAAATCCAGATCAATCTATGCACACATGCAAGTGGACAAATG 1001
Db 841 ATAGGTACCATGCCAGAA-----GTTCTCTACTAGAAATGCGCGTCA 885
QY 1002 GATTCAGTGCACCCGAGAAACTCATCAATATCAATTTGGAGATGATGATTTTGAGCAG 1061
Db 886 GAACCAAGTGAACCCCTGAAATTTCTCGCTTCTTTTGGCGAGGATGACCTCTTTGAA 945
QY 1062 AGTTCCAAAAGTGAATTAATCAGGAGGGATGATATGATGAAGATGAACCTGATGCCAA 1121
Db 946 CAAGGATTCATGAATAAACACGAGATG-----ATGATGGAATGAACCTGATTTAA 999
QY 1122 AGATGAAAAATGAAGTGA---AAATGAGGGTATCTCAGCCCTCAGCCCTGCAAGTGAACAGTG 1178
Db 1000 AGATGGAAGGTGAATATGAGAGTAAATGAACCAATGCTATCTCTGGAGTGAATGCTGTC 1059
QY 1179 AGAGAACCTAGAGTTGTTAGTTCAGACAAACAGTGCATTTGATATCTTCTGATGATGCTAT 1238
```

Db 1060 CGCGAACCAAGATTGTAGTCCAAACAACTAGTAGATATAGATATCTTTGATGATGTTAT 1119
 Qy 1239 AGTGGAGAAAATACCGGCGAAGAGTAGTGAAGGCAATCCAAATCCAGGAGTTACTAC 1298
 Db 1120 AGTGGAGAAAATAGTACGACAGAGTTGTAAAGGCAATCCAAATCCTAGGAGTACTAC 1179
 Qy 1299 AAGTGCACACACCCAGATCTCCAGTGCAGGAGCAGCTGGAAGAGCCTCACATGACCTA 1358
 Db 1180 AATGTGTACAAAGTTGGCTGTCCAGTGAGAAAGCAGCTTGAAGAGCATCTCATGATCTA 1239
 Qy 1359 AGGCTGTGATCAAACTTATGAGGGAAGCAACCAACCATGATGTTCTGAGCCCGTGGC 1418
 Db 1240 AGGCACTGATCACCACTTACGAGGAACACATAACCAACGATGTTCTGCCCCGCGAGC 1299
 Qy 1419 AGTGGCAGCATCTCTGTGAACAGACCAATGCGCAAAACAAATGCTTCAACACACACCAACT 1478
 Db 1300 AGTGGCAGTTATCTCTG-----CAGTGAACAGGCGCATCTGATACACACCACCACT 1347
 Qy 1479 GCAGCCACTTCCGTAAAGCTTTGCCAGTGTCCACCAAGTGACATTCCTTCCAGAAC 1538
 Db 1348 GCTCCTACGGTATAAGGCCAACACCA-----AATTACTTAAACCTCTTTCAGAAC 1398
 Qy 1539 CAAAGATCACAAAGCACCACAGAGGCAATCACCTTCCACCTTAGAGATGCTACAAAGT 1598
 Db 1399 CCAAGAGCACA--ACCTGCCAATGGTCAAGCACCAATTACCTAGAAATGCTGCAACGT 1455
 Qy 1599 CCAGGAAGTTTGGATCTCAGGTTTGGGAATCCAATGCAATCTTACGTGAACAGCAG 1558
 Db 1456 CCAGGAAGTACGAGTTCTCAGGTTTACAAACACATCCAATCTTACGGGATCAACCAA 1515
 Qy 1659 CAATATCTGCATGTTTCTCTCCAGGACCAAGGAGGCGCTAGAGATGACATGTTTC 1718
 Db 1516 AATCAACAGGCGCAGTGACAGTCTTCCACAGCCCAAGGACGAGCCTGATGTTGACTCGTTT 1575
 Qy 1719 CTGAGTCTCTACTATGCTGA 1739
 Db 1576 TTGACTCTTTTATGCTGA 1596

RESULT 9
 AB063575
 LOCUS
 DEFINITION Nicotiana tabacum NtWRKY-8 mRNA for WRKY DNA-binding protein,
 partial cds.
 ACCESSION AB063575
 VERSION AB063575.1 GI:14530684
 KEYWORDS
 SOURCE Nicotiana tabacum (common tobacco)
 ORGANISM Nicotiana tabacum
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 Asteridae; Lamiids; Solanales; Solanaceae; Nicotiana.
 1 (sites)
 Maeo, K., Hayashi, S., Kojima-Suzuki, H., Morikami, A. and Nakamura, K.
 Role of conserved residues of the WRKY domain in the DNA-binding
 activity of tobacco WRKY family proteins
 Unpublished
 2 (bases 1 to 1417)
 Maeo, K., Hayashi, S., Kojima-Suzuki, H., Morikami, A. and Nakamura, K.
 Direct Submission
 Submitted (21-JUN-2001) Kenichiro Maeo, Nagoya university, Graduate
 school of biocultural sciences; Chikusa-ku, Nagoya, Aichi
 464-8601, Japan (E-mail:maeo@agr.nagoya-u.ac.jp,
 Tel:81-052-789-4096, Fax:81-052-789-4095)
 Location/Qualifiers
 1. 1417
 /organism="Nicotiana tabacum"
 /mol_type="mRNA"
 /cultivar="Petit Havana SRI"
 /db_xref="taxon:4097"
 1. 1417
 /gene="NtWRKY-8"
 <2. 1138
 CDS
 gene

/gene="NtWRKY-8"
 /note="tobacco WRKY family proteins"
 /codon_start=1
 /product="WRKY DNA-binding protein"
 /protein_id="BAB61055.1"
 /db_xref="GI:14530684"
 /translation="KNYSDFSPQPOFASVPSQTNPLGKSNWYQSRKQNDENAN
 GTSLEQLKNGQSNQYKQSRSDGYNWRKYGQKVGKSNPRSYKCTPPNCTPK
 KVEELDDQITETIVYKGNHHPKPTQSRSSSLAIQPYNTQTNIEIPHQSTPENS
 ISFGDDHKSRSRGGDDPEEEDPSKEDPKRWKRESEGLSPKSGSTVREPRVVV
 QITSDILDDGTRWRKYQKVKGNPNRKYKCTSPCPYRKVHRASQDIRSVIT
 TTYGKHNDVPAARSAINRPVNTIYNNAIPIRPSVTQIPLPQSFTEMLHKP
 SNYGFSGVATSEDSYENQLDNNVPSRAKDEPRDDMFMETLLC"
 BASE COUNT 513 a 269 c 275 g 360 t
 ORIGIN
 Query Match 18.3%; Score 353.4; DB 8; Length 1417;
 Best Local Similarity 68.7%; Pred. No. 8.5e-75;
 Matches 571; Conservative 0; Mismatches 206; Indels 54; Gaps 4;
 Qy 684 CAATCTCAGACTTTTAGTAGAAGTCCAGATGATGGGTCAAAATTTGGAGGAAATATGCCAA 743
 Db 179 CAATATAAACAAGCAATCATCAAGATCAGAAGATGGGTACAAATTTGGAGGAAATATGGCAG 238
 Qy 744 AAACAAGTGAAGGGAAGTGAATAATCCAAGAAGTTATTACAAATGCACATACCCCAATTGC 803
 Db 239 AAACAAGTGAAGGTTAGTGAATAATCCAAGAAGTTACTATAAATGTACTTTCCCAATTTG 298
 Qy 804 CCTACAAGAAGAGTTGAGAGGCTTTAGATGGAACAAATTTACTGAGATAGTTTATAG 863
 Db 299 CCCACAAGAAAAAGTAGAGAGGTTTGTAGATGGGCAAAATTAATGAAATTTGTGTACAAG 358
 Qy 864 GGTACTCTAATACCATCTTAAGCTCAAAATCTAGGAGAACTCATCAAACTCTCTCT 923
 Db 359 GGTAATCAACAACCATCCAAGCGCACTCAGTCTCTAG-----GAGATCTCTCT 409
 Qy 924 CTTCGAATCCCTCATTTCAAATTCATCAGAACTGAAATCCCAGATCAATCAATATCAATTGGA 983
 Db 410 TTAGCCATCAACCTTTACAATACACAACTAATGAATCCAGATCATCAGT-----461
 Qy 984 CATGGAAGTGGACAAATGGAATTCAGTGCACCCAGAAACTCATCAATATCAATGGA 1043
 Db 462 -----CAACACCTGAGAAATTCGTCATTTTCTTGGG 493
 Qy 1044 GATGATGATTTTCAGCAGAGTTTCCCAA-----AGTGAATCAGAGGGGATGAATAT 1097
 Db 494 GATGATGATCATGAGAGAGTAGTCAAGAGGAGATGATTTCCATGAGAGAGAACCA 553
 Qy 1098 GATGAAGATGAACCTTGATGCCAAAAGATGAAAATTTGAAGGTGAAAATGAGGATATGTCA 1157
 Db 554 GACTCAAAAGAACCAAGACCCCAAAAGATGAAAAGAGAAAGCGAAAGTGAAGTTTATCA 613
 Qy 1158 GCCCCTGGAAGTGAACAGTGAAGAACCTTAGAGTTGTAGTTTCAGACCAACAGTGACATT 1217
 Db 614 GCACCAAGGAGCAGAACAGTGAAGAACCTTAGAGTTGTAGTTTCAAACTACAAGTGATTT 673
 Qy 1218 GATATCTTGATCATGCTATAGTGTGAGAAAATACGGCAGAAAAGTAGTGAAGGCGCAAT 1277
 Db 674 GATATTTAGATGATGATGATGATGAGGAGATGATGTTCAAAAGTAGTCAAAAGGAAT 733
 Qy 1278 CCAAAATCCAAGGAGTTACTACAAAGTGCACACCCAGGATGTCCAGTGAGGAAGACGCGT 1337
 Db 734 CCTAATCCAAGGAGTACTACAAATGCACAAAGTCCAGATGTCCAGTGAAGAAACATGTG 793
 Qy 1338 GAAAGAGCCTCAATGACCTTAAGGGTGTGATCACAATTTATGAGGGAAGACCAACCAT 1397
 Db 794 GAACGAGCATCAACAAGACATAAGATCAGTGATAACCACTTATGAAGGAAACCAACCAAC 853
 Qy 1398 GATGTTCTTCAGCCCGTGGCAGTG---GCAGGCAATTTCTGTGAACACACCAATGCCAAC 1454
 Db 854 GATGTTCCAGCAGCTAGGGGCGAGTGCATCAACCGACCTGTCCGACCAACCATTTACTTAC 913
 Qy 1455 AATGCTTCAAAACCAACCAACCACTGCAGCCACTTCGTAAAGGCTCTTTGCCA 1505

```
Db 914 AACAAATCCATCCCAATGAGCCTTCTGTACATCTCAAAATTCATTCGCCA 964
|||||
RESULT 10
AX412575 1539 bp DNA linear PAT 14-JUN-2002
LOCUS
DEFINITION Sequence 339 from Patent WO0222675.
ACCESSION AX412575
VERSION AX412575.1 GI:21445033
KEYWORDS
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1
Glazebrook,J., Wang,X., Dangl,J.L., Eulgem,T. and Zhu,T.
Plant genes, the expression of which are altered by pathogen
infection
Patent: WO 0222675-A 339 21-MAR-2002;
Syngenta Participations AG (CH); UNIVERSITY OF NORTH CAROLINA AT
CHAPEL HILL (US); Glazebrook, Jan (US); Wang, Xun (US); Dangl,
Jeffrey L. (US); Eulgem, Thomas (US)
LOCATION/Qualifiers
source
1. .1539
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
BASE COUNT 506 a 355 c 309 g 369 t
ORIGIN
Query Match 16.0%; Score 308.2; DB 6; Length 1539;
Best Local Similarity 67.1%; Pred. No. 7.7e-64;
Matches 493; Conservative 0; Mismatches 218; Indels 24; Gaps 3;
QY 710 AGATGATGGGTACAAATGGAGGAATATGCGCAAAACAAAGTGAAGGAAAGTGAAGATCC 769
DB 528 AGAGGATGGTTACAAATGGAGGAATATGCGCAAAACAAAGTGAAGGAAAGTGAAGATCC 587
QY 770 AAGAGTATTTACAAATGACATACCCCAATTCGCCCTTACAAAGAGAGGTTGAGAGTCC 829
DB 588 TCGGAGTTACTATAAGTGTACTTTCCCTTAATTTGCCAAGAGAGAAAGTGGAGAGATC 647
QY 830 TTTAGATGGCAAAATTAAGTGTACTTTATAGGCTTCTATAGGCTTCTATAGCCTTCTAGCCTCA 889
DB 648 TTTGGAAGGTCAGATCAGAGATTTGTGTATAAGGAAAGCCAAACATCTCTAAACCTCA 707
QY 890 AAATCTAGGAGAACTCATCAAACTCTCTCTCTTTCGCAATCCCTCATTCAAATTCAT 949
DB 708 GTCTACTAGAGATCTTCTCGTCTCTTCGACTTTTCATTCAGCTGTGTACATGCCAG 767
QY 950 CAGAACTGAATCC-----CAGATCAATCTCTATGCGGAGGATGAATATGATGAAGATGA 997
DB 768 TTTGGATCATAATCGTCAAGCTTCTCTGATCAGCTTAATTCCTCAATAGCTTTTCATCA 827
QY 998 AATGGATTGAGTCCCAACCCAGAA-----AACTCATCAATCAATTCGAGATGA 1048
DB 828 GTCTGATCTCTTTGGGATGCAACAAGAGGATAATCTACTTCTGATTTCTGTGGTGACGA 887
QY 1049 TGATTTTGAGCAGAGTTTCCCAAAAGTGAATCAGGAGGGGATGAATATGATGAAGATGA 1108
DB 888 TGAGTTGCAACAGGCTC---ATCGATTGTGACGAGAGAGAGAGATTTGTGGAGTGA 944
QY 1109 ACCTGATGCCAAAGATGGAAATTTGAAGTGAATTAAGGATGATGATGAGCCCTCGAAG 1168
DB 945 ACCTGAAGCAAGAGATGGAAAGGGGCAATGAAACAAATGGTGGGAATGGTGGTGAAG 1004
QY 1169 TAGAAGCTGAGAGAACTGAGTTGTAGTTTCAGACAAACAGTGCATATTCATCTTGA 1228
DB 1005 CAAGACATGAGAGAGCCGGAATCGTAGTGCAGACAAACAGGATGATATTCATCTTGA 1064
QY 1229 TGATGGCTATAGGTGGAGAAATACGGGCGAGAAAGTAGTAGTGAAGGGCAATCCAAATCCAAAG 1288
|||||
Db 1065 CGACGGTTACAGATGGAGAAATACGCCAGAAAGTCGTTAAGGGAACCCAAATCCAAG 1124
|||||
QY 1289 GAGTTACTCAAGTGCACACACCCAGCATGTCAGTGAAGACACGTCGGAAGAGCCTC 1348
|||||
DB 1125 AAGCTACTCAAGTGCACACACCATCGTTGTCTGAGTGAAGAAACATGTGGAGAGATC 1184
|||||
QY 1349 ACATGACCTAAGGGCTGTGATCAACATTTATGAGGAAAGCACAACCATGATCTTCTGTC 1408
|||||
DB 1185 ACACGACATGAGAGCAGTAATCAACACTACGAAAGGGAACACACACCATGTTCTCTGC 1244
|||||
QY 1409 AGCCCTGGCAGTGG 1423
|||||
DB 1245 AGCTCGTGGTAGCGG 1259
|||||
RESULT 11
AX412905 1539 bp DNA linear PAT 14-JUN-2002
LOCUS
DEFINITION Sequence 669 from Patent WO0222675.
ACCESSION AX412905
VERSION AX412905.1 GI:21445363
KEYWORDS
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.
1
Glazebrook,J., Wang,X., Dangl,J.L., Eulgem,T. and Zhu,T.
Plant genes, the expression of which are altered by pathogen
infection
Patent: WO 0222675-A 669 21-MAR-2002;
Syngenta Participations AG (CH); UNIVERSITY OF NORTH CAROLINA AT
CHAPEL HILL (US); Glazebrook, Jan (US); Wang, Xun (US); Dangl,
Jeffrey L. (US); Eulgem, Thomas (US)
LOCATION/Qualifiers
source
1. .1539
/organism="Arabidopsis thaliana"
/mol_type="genomic DNA"
/db_xref="taxon:3702"
BASE COUNT 506 a 355 c 309 g 369 t
ORIGIN
Query Match 16.0%; Score 308.2; DB 6; Length 1539;
Best Local Similarity 67.1%; Pred. No. 7.7e-64;
Matches 493; Conservative 0; Mismatches 218; Indels 24; Gaps 3;
QY 710 AGATGATGGGTACAAATGGAGGAATATGCGCAAAACAAAGTGAAGGAAAGTGAAGATCC 769
DB 528 AGAGGATGGTTACAAATGGAGGAATATGCGCAAAACAAAGTGAAGGAAAGTGAAGATCC 587
QY 770 AAGAGTATTTACAAATGACATACCCCAATTCGCCCTTACAAAGAGAGGTTGAGAGTCC 829
DB 588 TCGGAGTTACTATAAGTGTACTTTCCCTTAATTTGCCAAGAGAGAAAGTGGAGAGATC 647
QY 830 TTTAGATGGCAAAATTAAGTGTACTTTATAGGCTTCTATAGGCTTCTATAGCCTTCTAGCCTCA 889
DB 648 TTTGGAAGGTCAGATCAGAGATTTGTGTATAAGGAAAGCCAAACATCTCTAAACCTCA 707
QY 890 AAATCTAGGAGAACTCATCAAACTCTCTCTCTTTCGCAATCCCTCATTCAAATTCAT 949
DB 708 GTCTACTAGAGATCTTCTCGTCTCTTCGACTTTTCATTCAGCTGTGTACATGCCAG 767
QY 950 CAGAACTGAATCC-----CAGATCAATCTCTATGCGGAGGATGAATATGATGAAGATGA 997
DB 768 TTTGGATCATAATCGTCAAGCTTCTCTGATCAGCTTAATTCCTCAATAGCTTTTCATCA 827
QY 998 AATGGATTGAGTCCCAACCCAGAA-----AACTCATCAATCAATTCGAGATGA 1048
DB 828 GTCTGATCTCTTTGGGATGCAACAAGAGGATAATCTACTTCTGATTTCTGTGGTGACGA 887
QY 1049 TGATTTTGAGCAGAGTTTCCCAAAAGTGAATCAGGAGGGGATGAATATGATGAAGATGA 1108
|||||
```

888 TGATTCGAACAAGGCTC---ATCGATTGTCAGCAGAGCAAGAGATTGTGGAGTGA 944
 1109 ACCTGATGCCAAAGAGATGGAAATTTGAAGGTGAAATAGGGTATGTCAGCCCTCGAAG 1168
 945 ACCTGAGCAAGAGATGGAAAGGGGCAATGAAACAAATGGTGGGAATGGTGGGAAG 1004
 1169 TAGAACAGTAGAGAACCTAGAGTTGTAGTTTCAGACAAACAGTGACATTTGATTCCTTGA 1228
 1005 CAAGACAGTAGAGAGCGGAATCGTAGTGCAGACAAACGAGTGATATTCACATTCCTTGA 1064
 1229 TGATGGCTATAGTGGGAGAAATACGGCGCAGAAAGTAGTGAAGGGCAATCCAAATCCAAG 1288
 1065 CGACGGTTACAGTGGAGAAATACGGCCAGAAAGTCTGTAGGGAAACCCAAATCCAAG 1124
 1289 GAGTTACTACAGTGCAACCCAGGATGTCAGTGAAGAGCACACCAATCATGATGTTCTGTC 1348
 1125 AAGCTACTACAAGTGCAACCACTCGGTTGTCAGTGAGGAAACATGTGGAGAGGATC 1184
 1349 ACATGACCTAAGGCTGTGATCAACCTTATGAGGGAAGCACCAACCATGATGTTCTGTC 1408
 1185 ACAGACATGAGAGCAGTAATCAACCTAGCAAGGGAACACCAACCATGATGTTCTGTC 1244
 1409 AGCCCGTGGCAGTGG 1423
 1245 AGCTCGTGGTAGCGG 1259

RESULT 12
 AX506109 1539 bp DNA linear PAT 27-SEP-2002
 LOCUS
 DEFINITION Sequence 804 from Patent WO0216655.
 AX506109
 ACCESSION
 VERSION AX506109.1 GI:23387346
 KEYWORDS
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi
 1
 Harper, J.F., Kreps, J.J., Wang, X. and Zhu, T.
 Stress-regulated genes of plants, transgenic plants containing
 same, and methods of use
 Patent: WO 0216655-A 804 28-FEB-2002;
 The Scripps Research Institute (US) ; Syngenta Participations AG
 (CH)

FEATURES
 source
 1. .1539
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /db_xref="taxon:3702"
 BASE COUNT 506 a 355 c 309 g 369 t
 ORIGIN
 Query Match 16.0%; Score 308.2; DB 6; Length 1539;
 Best Local Similarity 67.1%; Pred. No. 7.7e-64;
 Matches 493; Conservative 0; Mismatches 218; Indels 24; Gaps 3;
 710 AGATGATGGGTACAAATGGAGGAATATGGCCAAACCAAGTGAAGGAGTGAATCC 769
 528 AGAGATGGTTACAAATGGAGGAATATGGCCAAACCAAGTGAAGGAGTGAATCC 587
 770 AAGAAGTTATTACAAATGCACATACCCCAATTCGCCCTACAAGAGAGGTTGAGAGTCC 829
 588 TCGAGTTACTATAAGTGTACTTTCCTTAATTTGTCACGAGGAAGAAAGTGGAGAGATC 647
 830 TTTAGATGGCAAAATTAATCTAGATAGTTTATAAGGGTACTCATTAACCATCTTAAGCCTCA 889
 648 TTTGGAAGGTGAGATCAAGAGATTTGTGTATAAAGGAAGCCACCAACCATCTCTAAACCTCA 707
 890 AAATACTAGGAGAACTCATCAAACTCTCTCTCTCTGCAATCCTCTCATTCATTCATTCAT 949

708 GTCTACTAGAGATCTTCTTCTGCTTTCTTCGACTTTTCTATTCAGCTGTGTACAAATCCAG 767
 950 CAGAACTGAATCC-----CAGATCAATCTATGCCACACATGGAAGTGGACA 997
 768 TTTGGATCAATAATCGTCAAGCTTCTTCGATCAGCCTAATTCATAATAGCTTTTCATCA 827
 998 AATGGATTTCAGCTGCCACCCAGAA-----AATCATCAATATCAATTTGGAGATGA 1048
 828 GTCTGATTCCTTTGGGATGCCACAGAGGATAATACTACTTCTGATTTCTGTTGGTGACGA 887
 1049 TGATTTTGAGCAGAGTTCCTCAAAAGTGTAAATCAGAGGGGATGAATATGATGAAGATGA 1108
 888 TGAGTTTCGAACACAGGCTC---ATCGATTGTCAGCAGACGACGAAGAAGATTGTGGGAGTGA 944
 1109 ACTGATGCCAAAGATGCAAAATTTGAAGGTGAAATGAGGTATGTCAGCCCTCGAAG 1168
 945 ACCTGAAGCAAGAGATGGAAAGGGGCAATGAACAAATGGTGGGAATGGTGGTGAAG 1004
 1169 TAGAACAGTGAGAGAACCTTAGAGTTGTAGTTTCAGACAAACCAAGTGAATTCATTCCTTGA 1228
 1005 CAGACAGTGAGAGAGCCCGAATCGTAGTGCAGACACACGAGTGATATTGACATTCCTTGA 1064
 1229 TGATGCTATAGTGGAGAAATACGGGCGAGAAAGTAGTGAAGGGCAATCCAAATCCAAG 1288
 1065 CGACGGTTACAGATGGAGAAATACGGCCAGAAAGTCGTTAAGGGAAACCCAAATCCAAG 1124
 1289 GAGTTACTACAGTGCAACCCAGGATGTCAGTGAAGAGCACGTCGGAAGAGCCTC 1348
 1125 AAGCTACTACAGTGCAACCACTCGGTTGTCAGTGAGGAAACATGTGGAGAGGATC 1184
 1349 ACATGACCTAAGGCTGTGATCAACCTTATGAGGGAAGCACCAACCATGATGTTCTGTC 1408
 1185 ACAGACATGAGAGCAGTAATCAACCTACGAGGGAACACACACCATGATGTTCTGTC 1244
 1409 AGCCCGTGGCAGTGG 1423
 1245 AGCTCGTGGTAGCGG 1259

RESULT 13
 AX651760 1539 bp DNA linear PAT 22-MAR-2003
 LOCUS
 DEFINITION Sequence 606 from Patent WO03000898.
 AX651760
 ACCESSION
 VERSION AX651760.1 GI:29154578
 KEYWORDS
 SOURCE Arabidopsis thaliana (thale cress)
 ORGANISM Arabidopsis thaliana
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi
 1
 Chang, H.S., Chen, W., Cooper, B., Glazebrook, J., Goff, S.A., Hou, Y.M.,
 Katagiri, F., Quan, S., Tao, Y., Whitham, S., Xie, Z., Zhu, T. and Zou, G.
 Plant genes involved in defense against pathogens
 Patent: WO 03000898-A 606 03-JAN-2003;
 Syngenta Participations AG (CH)
 FEATURES
 source
 1. .1539
 /organism="Arabidopsis thaliana"
 /mol_type="genomic DNA"
 /db_xref="taxon:3702"
 BASE COUNT 506 a 355 c 309 g 369 t
 ORIGIN
 Query Match 16.0%; Score 308.2; DB 6; Length 1539;
 Best Local Similarity 67.1%; Pred. No. 7.7e-64;
 Matches 493; Conservative 0; Mismatches 218; Indels 24; Gaps 3;
 710 AGATGATGGGTACAAATGGAGGAATATGGCCAAACCAAGTGAAGGAGTGAATCC 769
 528 AGAGATGGTTACAAATGGAGGAATATGGCCAAACCAAGTGAAGGAGTGAATCC 587
 770 AAGAAGTTATTACAAATGCACATACCCCAATTCGCCCTACAAGAGAGGTTGAGAGTCC 829
 588 TCGAGTTACTATAAGTGTACTTTCCTTAATTTGTCACGAGGAAGAAAGTGGAGAGATC 647
 830 TTTAGATGGCAAAATTAATCTAGATAGTTTATAAGGGTACTCATTAACCATCTTAAGCCTCA 889
 648 TTTGGAAGGTGAGATCAAGAGATTTGTGTATAAAGGAAGCCACCAACCATCTCTAAACCTCA 707
 890 AAATACTAGGAGAACTCATCAAACTCTCTCTCTCTGCAATCCTCTCATTCATTCATTCAT 949

Qy 1289 GAGTTACTACAGTGCACACACCCAGGATGTCCAGTGGAGGACACGTGGAAGAGCGCTC 1348
Db 1125 AAGTACTACAGTGCACACACCCAGGATGTCCAGTGGAGGACACGTGGAAGAGCGCTC 1184
Qy 1349 ACATGACCTAAGGCTGTGATCAACAATTATAGGAGGAAAGACACACCATGATGTTCTGC 1408
Db 1185 ACACGACATGAGAGCAGTAAATCAACACTAGAGGAGGAAACACACCATGATGTTCTGC 1244
Qy 1409 AGCCCGTGGCAGTGG 1423
Db 1245 AGCTCGTGGTAGCGG 1259

RESULT 15
AX654743
LOCUS AX654743 1668 bp DNA linear PAT 22-MAR-2003
DEFINITION Sequence 4613 from Patent WO0300898.
ACCESSION AX654743
VERSION AX654743.1 GI:29157557
KEYWORDS
SOURCE Oryza sativa
ORGANISM Oryza sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzeae; Oryza.

REFERENCE
1 Chang, H.S., Chen, W., Cooper, B., Glazebrook, J., Goff, S.A., Hou, Y.M.,
Katagiri, F., Quan, S., Tao, Y., Whitman, S., Xie, Z., Zhu, T. and Zou, G.
Plant genes involved in defense against pathogens
Patent: WO 0300898-A 4613 03-JAN-2003;
Syngenta Participations AG (CH)
Location/Qualifiers
1..1668
/organism="Oryza sativa"
/mol_type="genomic DNA"
/db_xref="taxon:4530"
BASE COUNT 326 a 554 c 574 g 212 t 2 others
ORIGIN

Query Match 14.4%; Score 277.2; DB 6; Length 1668;
Best Local Similarity 54.8%; Pred. No. 2.5e-56;
Matches 696; Conservative 1; Mismatches 534; Indels 39; Gaps 6;

Qy 174 GGCTCTGGTGTCCCAATTCAGTCCACACACACCTTCTGCGCTCTCTCTCTCTCTCTCTCT 233
Db 115 GCGGAGGTGTCTCAGGTACAGGCCATGACCCGCTTCTGCTGCGCTCTCTCTCTCTCTCTCT 174
Qy 234 CCCATTTCT 293
Db 175 CCGGTGTGCGGTCT 234
Qy 294 GACTCGCGGTCT 353
Db 235 GGCTCGCGGTCT 291
Qy 354 GTTCTCAGAGCTTCAATTCGAAGAGAGTTCAGGGGGATCAGCAATTTGTCAAGGAA 413
Db 292 GCGTGGAGACACTTTGACTGAGGCGGAGAGGTGGCGGCGGCGGAGCGCGGATCAAGGC 351
Qy 414 GAAGACAAAAGCTTCTCAAAATTTCTCTTTTCCAAACCCGATCAGGACCTCTCTCTCTCTCT 473
Db 352 GGCAGGACGAGCAGAGGAATTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 411
Qy 474 ACAGCAACATACAGTCTTCAAAATGTCAAGTTCAAAACACACAG---CCATGAGATTTT 530
Db 412 GAGGAGGCGGTGCGAAACAGACACTTTCCAGCCACCGGTCTCCACCGGCGCCCACTGGGGAC 471
Qy 531 CAGGAGGCGGACGAAACAGATTAATTTTCTCTCAGAAAGGTATGATCAAACTGAA---587
Db 472 GAGGACATACAGAGTTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 531
Qy 588 ---AAGTCT 644

Db 532 ATGGACCGCGGTGCCAAACCGCGCGAGCTTTCGGTGGCGCGCGCTTTCGAGCGACCTCGTCG 591
Qy 645 AGCAATGGGTGTTTCAATTCGATTTATGCAATTTACCCCCACAAATCTCAGAGCTTTAAAGTAGA 704
Db 592 GAGATGCGCCACAGGTGCAGGGTGGCGGGGTACAGCCAGCCGCGAGTCGCAGAGCGCG 651
Qy 705 AGGTCAAGATGATGGGTACAAATTTGGAGGAAATATGGCCAAAAAACAAGTGAAGGAAGTGA 764
Db 652 TGTCGGACGACGGCTACAACTGGCGCAAGTAGTCGGGCAAGCAGGTGAAGGGAGCGAG 711
Qy 765 AATCCAAGAGTATTATCAAAATGCACATACCCCAATTCCTTCCCTTACAAAGAGAGGTTGAG 824
Db 712 AACCCCGCAGCTACTACAACTGCACTTCCCGAATTCGCCCAAGCAAGAGAGGTTGAG 771
Qy 825 AGTCTTTTATAGTGCACAAATTTACTGATAGTAGTTTATAAGGCTACTCATCAACCATCTCAAG 884
Db 772 CGGTCTCTCGACCGCCAGATCACCGAGATCTGTACAGGGGACGCAACACCGCCAG 831
Qy 885 CCTCAAAATATCTAGGAGAACTCATCAAACTCCTCTCTCTCTCTCTCTCTCTCTCTCTCT 944
Db 832 CCGCAGAACACGCGCAGGAATCTCCGCTCGTGGCGCGCAGGTCTCTGC-----880
Qy 945 TCATCAGAACTGAAATCCAGATCAATCTTATGCCACACATGGAAGTGCACAAATGGAT 1004
Db 881 -AGAGCGCGCGGACATGTCCGAGCATCTCTTCG-----GGGCGATGTCCGCG 927
Qy 1065 TCCCAAAAGT---GTAAATCAGAGGGGATGATATGATGAAGATGAACCTGATGCCAA 1121
Db 988 TCCCTCTCGGCGCGGAAACGCGCGCGCAGAGTTCGACGACGACGAGCGGATTCGAAG 1047
Qy 1122 AGATGAAATTTGAAGTGAATATGAGGTATGTAGCCCTTGGAAAGTGAACAGTGA 1181
Db 1048 AGATGAGGAAAGACGCTGACGCGGAGGGATCTCCATGGCTGGCAACCGGACGGTCCG 1107
Qy 1182 GAACCTAGAGTGTAGTTCAGACCAACAGTGCATTTGATATCTCTGATGCTATAGG 1241
Db 1108 GAACCGAGGTGTCTCTCAACCATGAGGACATCGACATCTCTCGACGCGCTACCGC 1167
Qy 1242 TGGAGAAAATACCGGCGAGAAAGTAGTGAAGGGCAATCCAAATCCAAAGGAGTTACTACAAG 1301
Db 1168 TGGAGAAAGTACCGGCGAGAAAGTGTGTGAAGGGCAACCCGAAACCAAGGAGCTACTACAAG 1227
Qy 1302 TGACACACCCAGATGTCCAGTGAAGACAGTGAAGAGCCTCAGATGACCTAAGG 1361
Db 1228 TGACACACCGCGGTTCCTCTGCGGAGACAGTGGAGCGCGCTCCACGACCTGCGC 1287
Qy 1362 GCTGTGATCAACTTTATGAGGAAAGCAACCATGATGTTCTCTGAGGCCCTGGCAGT 1421
Db 1288 GCGTGTATCAACCTTACGAGGGGAAACACACGACGTGCGCGCGCGCGCGCGCGGAGC 1347
Qy 1422 GCGAGCCATT 1431
Db 1348 GCGCGCTTT 1357

Search completed: January 20, 2004, 18:26:01
Job time : 7205 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: January 20, 2004, 16:14:39 ; Search time 4188 Seconds
(without alignments)
11188.880 Million cell updates/sec

Title: US-09-890-811b-9

Perfect score: 1928

Sequence: 1 gcacgagtctatgcgcctc.....tcaaaaaaaaaaaaaaaaaa 1928

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- EST.*
- 1: em_estba.*
 - 2: em_esthum.*
 - 3: em_estin.*
 - 4: em_estnu.*
 - 5: em_estov.*
 - 6: em_estpl.*
 - 7: em_estro.*
 - 8: em_hc.*
 - 9: gb_est1.*
 - 10: gb_est2.*
 - 11: gb_hc.*
 - 12: gb_est3.*
 - 13: gb_est4.*
 - 14: gb_est5.*
 - 15: em_estfun.*
 - 16: em_estom.*
 - 17: em_gss_hum.*
 - 18: em_gss_inv.*
 - 19: em_gss_pin.*
 - 20: em_gss_vrt.*
 - 21: em_gss_fun.*
 - 22: em_gss_mam.*
 - 23: em_gss_mus.*
 - 24: em_gss_pro.*
 - 25: em_gss_rod.*
 - 26: em_gss_plg.*
 - 27: em_gss_vrl.*
 - 28: gb_gss1.*
 - 29: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	610.2	31.6	673	9	AW133440
c 2	603.4	31.3	662	13	BU550574
c 3	597.8	31.0	757	10	BE820590
c 4	583.8	30.3	746	10	BE658511

c	5	583.6	30.3	749	10	BE657634
	6	579	30.0	603	13	BQ295846
	7	568.8	29.5	572	13	BQ080806
	8	553.6	28.7	568	13	BU091659
	9	528.4	27.4	531	12	BI975061
	10	515	26.7	585	12	BM519918
	11	505.4	26.2	507	12	BM143621
	12	499.4	25.9	509	13	BU578260
	13	485.4	25.2	772	14	CB894319
	14	484.4	25.1	570	14	CD042161
	15	471.8	24.5	518	12	BM187988
	16	462.4	24.0	577	12	BM568389
c	17	458.4	23.8	559	12	BG840033
c	18	453.2	23.5	657	12	BI967766
	19	449.4	23.3	451	13	BU762115
	20	434	22.5	435	10	BE803423
	21	422.4	21.9	424	12	BM890972
	22	413.4	21.4	635	13	BU645573
	23	411	21.3	532	12	BM188894
	24	407.4	21.1	440	10	BG654191
	25	405.8	21.0	480	12	BM520930
	26	404.8	21.0	408	13	BU764331
	27	404.4	21.0	406	10	BG508622
	28	395.2	20.5	667	14	CD007418
	29	386.2	20.0	652	14	CD008237
	30	384.8	20.0	408	13	BU764518
	31	381.8	19.8	646	14	CD008324
	32	381.6	19.8	650	14	CD008395
	33	380.8	19.8	577	14	CA937181
	34	379	19.7	640	13	BU043758
	35	374.6	19.4	607	14	CA821401
	36	372.4	19.3	386	10	BG046426
	37	370	19.2	370	10	BE802905
	38	367.8	19.1	425	12	BM177021
	39	365.4	19.0	444	9	AW704231
	40	362.6	18.8	442	9	AW432500
	41	359.2	18.6	717	12	BI923269
	42	358.4	18.6	453	12	BM520345
	43	357.4	18.5	409	12	BI321844
	44	355	18.4	637	9	AI771129
	45	351.2	18.2	420	9	AU251455

ALIGNMENTS

RESULT 1
AW133440

LOCUS
DEFINITION

AW133440 673 bp mRNA linear EST 02-DEC-2001
se19a06.y1 Gm-cl015 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
Gm-cl015-1643 5' similar to TR:Q40090 Q40090 SPPI PROTEIN. ;, mRNA
sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

AW133440
AW133440.1 GI:6135047
EST.
Glycine max (soybean)
Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.

REFERENCE
AUTHORS

1 (bases 1 to 673)
Shoemaker, R., Keim, P., Vodkin, L., Erpelting, J., Corvill, V., Khanna
A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,
Wyllie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers
Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk
R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann
R., Waterston, R. and Wilson, R.
Public Soybean EST Project

TITLE
JOURNAL
COMMENT

Unpublished
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810

Email: est@watson.wustl.edu
 This clone is available through: ResGen, Invitrogen Corp. 2130
 South Memorial Parkway Huntsville, AL 35801 For further information
 call: (800)-533-4363 or contact via email: ccu@resgen.com
 Insert Length: 1363 Std Error: 0.00
 Seq primer: -40RP from Gibco
 High quality sequence stop: 416.
 Location/Qualifiers

FEATURES

source

1. .673
 /organism="Glycine max"
 /mol_type="mRNA"
 /db_xref="taxon:3847"
 /clone="GENOME SYSTEMS CLONE ID: Gm-cl015-1643"
 /tissue_type="Mature flowers, field grown plants"
 /lab_host="XL10-Gold"
 /clone_lib="Gm-cl015"
 /note="vector: pBluescript II XR; Site 1: EcoRI; Site 2:
 XhoI; This cDNA library was constructed from mRNA isolated
 from mature flowers of field grown plants. The cDNA
 library was prepared using the Stratagene pBluescript II
 XR cDNA library construction kit. Complementary DNA was
 synthesized from mRNA using a primer consisting of a poly
 (dT) sequence with a XhoI restriction site. EcoRI adapters
 were ligated to the blunt-ended cDNA fragments followed by
 XhoI digestion. The cDNA fragments were directionally
 cloned into the EcoRI-XhoI restriction site of the
 pBluescript vector. The ligated cDNA fragments were
 transformed into XL10-Gold host cells. This library was
 constructed by Dr. Randy Shoemaker and Dr. John
 Erpelnding."
 BASE COUNT 234 a 145 c 150 g 143 t 1 others
 ORIGIN

Query Match 31.68; Score 610.2; DB 9; Length 673;
 Best Local Similarity 95.58; Pred. No. 2e-97;
 Matches 638; Conservative 0; Mismatches 29; Indels 1; Gaps 1;
 QY 808 CAAAGAGAAGGTTGAGAGGCTTTAGATGGACAAATTTACTGAGATAGTTTATAAGGGTA 867
 DB 6 CAAAGAGAAGGTTGAGAGGCTTTAGATGGACAAATTTACTGAGATAGTTTATAAGGGTA 65
 QY 868 CTCATACCATCTTAAGCCTCAAAATACTAGGAGAAATCATCAAACTCTCTCTCTTG 927
 DB 66 CTCATACCATCTTAAGCCTCAAAATACTAGGAGAAATCATCAAACTCTCTCTCTTG 125
 QY 928 CAATCCCTCATTCATCAATCCATCAGAACTGAAATCCAGATCAATCTATGCCACATG 987
 DB 126 CAATCCCTCATTCATCAATCCATCAGAACTGAAATCCAGATCAATCTATGCCACATG 185
 QY 988 GAAGTGACAAATGGATTCAGTCCACCCAGACAAATCTCATCAATATCAATTTGGAGATG 1047
 DB 186 GAAGTGACAAATGGATTCAGTCCACCCAGACAAATCTCATCAATATCAATTTGGAGATG 245
 QY 1048 ATGATTTTGACAGAGTTTCCAAAGTGTAATCAGAGGGGATGAATATGATGAAGATG 1107
 DB 246 ATGATTTTGACAGAGTTTCCAAAGTGTAATCAGAGGGGATGAATATGATGAAGATG 305
 QY 1108 AACCTGATGCCAAAGATGGAATTCAGAGTGAATATGAGGATGTCAGCCCTCGAA 1167
 DB 306 AACCTGATGCCAAAGATGGAATTCAGAGTGAATATGAGGATGTCAGCCCTCGAA 365
 QY 1168 GTAGACAGTGAGAGAACTTAGTGTAGTTTCAGACAAACAGTGACATGATATCTCTTG 1227
 DB 366 GTAGACAGTGAGAGAACTTAGTGTAGTTTCAGACAAACAGTGACATGATATCTCTTG 425
 QY 1228 ATGATGCTATAGTGGAGAAATATACGGGCGAAGATGATGAAGGCAATCCAAATCCAA 1287
 DB 426 ATGATGCTATAGTGGAGAAATATACGGGCGAAGATGATGAAGGCAATCCAAATCCAA 485
 QY 1288 GGAGTTACTACAAGTGACACACCCAGATGTCCTCAGTGAGGAACGACGTGGAAAGAGCCT 1347

DB 486 GGAGTTACTACAAGTGACACACCCAGATGTCCTCAGTGAGGAACGCTGGAAAGAGCCT 545
 QY 1348 CACATGACCTTAAGGCTGTGATCACAACTATGAGGGAAGACACACCATGATGTTCTCTG 1407
 DB 546 CACATGACCTTAAGGCTGTGATCACAACTATGATGGAAGACACACCATGATGTTCTCTG 605
 QY 1408 CAGCCCGTGGCAGTGGCAG-CCATTCTGTGAACAGACCAATGCCAAACAATGTTCAAAC 1466
 DB 606 CCACCCCGTGGCAGTGGCAGCCCACTTGTACAGACCAATGCCACCAATGTTCAAAC 665
 QY 1467 CACACCAA 1474
 DB 666 CACAACCA 673

RESULT 2

BU550574/c

LOCUS

DEFINITION

BU550574

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Glycine max (soybean)

Glycine max

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 ; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Glycine.
 1 (bases 1 to 662)
 Vodkin,L., Shoemaker,R., Keim,P., Retzel,E., Khanna,A., Shealy,R.,
 Clough,S., Thibaud-Nissen,F., Coryell,V., Erpelnding,J., Raph,C.,
 Shoop,E., Scromvik,M., Schweitzer,P., Gong,G. and Liu,L.
 A Functional Genomics Program for Soybean (NSF 9872565) (2002)
 Unpublished
 Other ESTs: B1893195 corresponding to Gm-cl068-3332 (5')
 Contact: Vodkin, L.O., PI, A Functional Genomics Program for
 Soybean (NSF 9872565)
 Lewin, H. A., Director, Keck Center for Comparative and Functional
 Genomics
 University of Illinois
 Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
 Tel: (217) 244-6147
 Fax: (217) 333-4582
 Email: l-vodkin@uiuc.edu
 Insert Length: 662 Std Error: 0.00
 Plate: GM880021A20 row: G column: 03
 Seq primer: 5'-TTTTTTTTTTTTTTTTT(A/C/G)-3'
 High quality sequence stop: 662.
 Location/Qualifiers

FEATURES

source

1. .662
 /organism="Glycine max"
 /mol_type="mRNA"
 /db_xref="taxon:3847"
 /clone="Gm-r1088-7974"
 /clone_lib="Gm-r1088"
 /note="The library Gm-r1088 is a sequence-driven, reracked
 set of 9,216 low redundancy clones selected from cDNA
 libraries from various tissues and stages of development
 of soybean that consists of 2,706 cDNAs from germinating
 cotyledons (source library Gm-cl027); 1,355 cDNAs from
 immature seed coats (libraries Gm-cl019 and Gm-cl023); 917
 cDNAs from tissue culture derived somatic embryos (source
 libraries Gm-cl036 and Gm-cl075); 3,275 cDNAs from
 germinating seedlings, shoot tips, or leaves exposed to
 various stresses (source libraries Gm-cl064, Gm-cl065,
 Gm-cl066, and Gm-cl067; and Gm-cl068); and 963 cDNAs from
 young leaves exposed to bacterial and fungal pathogens
 (source libraries Gm-cl072, Gm-cl073, and Gm-cl074). The
 5' ESTs of the source clones from the different libraries
 was used to select singletons, or a representative of each
 contig, which were reracked to form library Gm-r1088 and

the cDNA clones of the reracked Gm-r1088 library were then sequenced at the 3' end. The unigene selection and 3' sequencing was funded by NSF Plant Genome project #9872565 (http://soybeanomics.cropsci.uiuc.edu/) as part of creation of a low redundancy soybean cDNA set. The source cDNA libraries were constructed by the laboratories of Lila Vodkin, University of Illinois, Randy Shoemaker, Iowa State University, and Paul Keim, Northern Arizona University as part of the Public EST project, http://129.186.26.94/soybeanest.html. The contig analysis to select unique genes was performed by the laboratory of Ernest Retzel, Center for Computational Genomics and Bioinformatics, University of Minnesota. Sequencing were conducted by services of the University of Illinois Keck Center for Comparative and Functional Genomics http://www.biotech.uiuc.edu/keck.htm. Note: The corresponding 5' EST from each clone in the Gm-r1088 library is listed in the 'OTHER EST' field. The detailed information on the source library for each clone can also be obtained by referring to the Genome Systems clone ID of the original cDNA library that is also listed under 'OTHER EST'."

BASE COUNT	179 a	145 c	146 g	192 t	
ORIGIN					
Query Match	31.3%; Score 603.4; DB 13; Length 662;				
Best Local Similarity	99.8%; Pred. No. 3.2e-96;				
Matches	604; Conservative	0; Mismatches	1; Indels	0; Gaps	0;
QY	1287	AGGAGTTACTACAAAGTGCACACCCAGGATGTCAGTGGAGGAGCAGTGGAAAGGCC	1346		
DB	605	AGGAGTTACTACAAAGTGCACACCCAGGATGTCAGTGGAGGAGCAGTGGAAAGGCC	546		
QY	1347	TGACATGACCTAAGGCTGTGATCAACTTATGAGGGAAGCACAACCATGATGTCCT	1406		
DB	545	TCACATGACCTAAGGCTGTGATCAACTTATGAGGGAAGCACAACCATGATGTCCT	486		
QY	1407	GCAGCCGCTGGCAGTGGCAGCACTTCTGTAACAGACCAATGCAACCAATGCTTCAAC	1466		
DB	485	GCAGCCGCTGGCAGTGGCAGCACTTCTGTAACAGACCAATGCAACCAATGCTTCAAC	426		
QY	1467	CACACCAACTCGACCACTTCCGTAAGGCTCTTCCAGTGATCCACCAAGTGACAAT	1526		
DB	425	CACACCAACTCGACCACTTCCGTAAGGCTCTTCCAGTGATCCACCAAGTGACAAT	366		
QY	1527	TCCCTTCAGAACCAAGATCAAGCACCAGAGAGGCAATCACCCTTCCCTTAGAG	1586		
DB	365	TCCCTTCAGAACCAAGATCAAGCACCAGAGAGGCAATCACCCTTCCCTTAGAG	306		
QY	1587	ATGCTACAAAGTCCAGGAAGTTTGGATTCTCAGGTTTGGGAATCCAATGCAATCTTAC	1646		
DB	305	ATGCTACAAAGTCCAGGAAGTTTGGATTCTCAGGTTTGGGAATCCAATGCAATCTTAC	246		
QY	1647	GTGAACCCAGCAGCAACTATCTGACAAATGTTTCTCTCCAGGACCAAGGAGGAGCCTAG	1706		
DB	245	GTGAACCCAGCAGCAACTATCTGACAAATGTTTCTCTCCAGGACCAAGGAGGAGCCTAG	186		
QY	1707	GATGACATGTTCTTGATGTTCTACTATGCTGAAGGAATTTTTTTTCCCTTTGGTA	1766		
DB	185	GATGACATGTTCTTGATGTTCTACTATGCTGAAGGAATTTTTTTTCCCTTTGGTA	126		
QY	1767	GCTATGGAAGTTGGAAATTTTGGAGTGGGGGACTAGGATTTATTTGGACAAATAGGTT	1826		
DB	125	GCTATGGAAGTTGGAAATTTTGGAGTGGGGGACTAGGATTTATTTGGACAAATAGGTT	66		
QY	1827	CCATTGCAATTTATGCAATTTTGGTTGTTTGTGTTAAATTTATACAGCCACAGGAT	1886		
DB	65	CCATTGCAATTTATGCAATTTTGGTTGTTTGTGTTAAATTTATACAGCCACAGGAT	6		
QY	1887	TGGTA	1891		
DB	5	TGGTA	1		

RESULT 3
BE820590/c
LOCUS
DEFINITION

BE820590
GM700012B20812 Gm-r1070 Glycine max cDNA clone Gm-r1070-4704 3',
mRNA sequence.

ACCESSION
VERSION
KEYWORDS

SOURCE
ORGANISM

BE820590.1 GI:10252824
EST.
Glycine max (soybean)

Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

1 (bases 1 to 757)
Vodkin,L., Keim,P., Shoemaker,R., Retzel,E., Khanna,A., Coryell,V.,
Erpeliding,J., Raph,C., Shoop,E., Fardinas,J., Liu,L. and Lewin,H.
A Functional Genomics Program for Soybean (NSF 9872565)
Unpublished

Other ESTs: AM396234 corresponding to Gm-cl016-5873 (5')
Contact: Vodkin, L.O., PI, A Functional Genomics Program for
Soybean (NSF 9872565)
Lewin, H. A., Director, Keck Center for Comparative and Functional
Genomics

University of Illinois
Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
Tel: (217) 244-6147
Fax: (217) 333-4582

Email: l-vodkin@uiuc.edu

This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134. For further information
call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or info@genome
systems.com web site: www.genomesystems.com

Seq primer: 5'-TTTTTTTCTTTTCTTTTCTTTT(A/C/G)-3'.

FEATURES
source

1..757
Location/Qualifiers
/organism="Glycine max"
/mol_type="mRNA"
/db_xref="taxon:3847"
/clone="Gm-r1070-4704"
/clone_lib="Gm-r1070"
/note="The library Gm-r1070 is a sequence-driven, reracked
set of 9,216 clones selected from cDNA libraries from
various tissues and stages of development of soybean that
represent 2,639 sequences from immature cotyledons, 1,770
from immature seed coats, 3,938 from flowers, and 869
from young pods. The 5' ESTs of the source clones from
the different libraries was used to select singletons, or
a representative of each contig, which were reracked to
form library Gm-r1070. The cDNA clones of the reracked
Gm-r1070 library were then sequenced at the 3' end. The
contig analysis to select unique genes was performed by
the laboratory of Ernest Retzel, Center for Computational
Genomics and Bioinformatics, University of Minnesota,
http://www.cbc.umn.edu/ResearchProjects/Soybean/index.html
Reracking was performed by Genome Systems, St. Louis,
http://www.genomesystems.com, and 3' sequencing by the
Keck Center for Comparative and Functional Genomics,
University of Illinois,
http://www.life.uiuc.edu/biotech/keck.html. Note: The
corresponding 5' EST from each clone in the Gm-r1070
library is listed in the 'OTHER EST' field. The detailed
information on the source library for each clone can also
be obtained by referring to the Genome Systems clone ID of
the original cDNA library that is also listed under
'OTHER EST'."

BASE COUNT 229 a 154 c 147 g 200 t 27 others

Query Match 31.0%; Score 597.8; DB 10; Length 757;

Best Local Similarity	95.5%;	Pred. No. 2.9e-95;
Matches 611;	Conservative	0; Mismatches 27; Indels 2; Gaps 1;

1277	Qy	TCCAAATCCAAAGAGTACTTACAAAGTGCACACCCAGAGATGTCAGTGGAGGACACGT	1333
757	Db	TCCAAATCCAAAGAGTACTTACAAAGTGCACACCCAGAGATGTCAGTGGAGGACACGT	698
1337	Qy	GGAAAGAGCCTCACATGACCTTAAGGGCTGTGATCAACAATTTATGAGGAAAGACACAACCA	1396
697	Db	NNNANGAGCCTCACATGANNNAAGNCTGTGATCAACAATTTATGAGGAAAGCANNNNCA	638
1397	Qy	TGATGTTCTTCGAGCCGCTGGCGAGTGGCAGCCCATTTCTGTGAAACAGACCAATGCCAAACAA	1456
637	Db	TGATGTTCTTCGAGCCGCTGGCGAGTNNACGCCAATTTCTGTTAAACAGACCAATGCCAAACAA	578
1457	Qy	TGCTTCAAAACCCACCAACACTGCAGCCACTTCGGTAAAGGCTCTTGCAGTGTATCCACCA	1516
577	Db	TGCTTCAAAACCCACCAACACTGCAGCCACTTCGGTAAAGGCTCTTGCAGTGTATCCACCA	518
1517	Qy	AAGTGACAAATTCCTTTTCAGAACCAAAAGATCACAAAGCACACGAGAGGGCAATCACCCCTT	1576
517	Db	AAGTGACAAATTCCTTTTCAGAACCAAAAGATCACAAAGCACACGAGAGGGCAATCACCCCTT	458
1577	Qy	CACCTTAGAGATGCTACAAAGTCCAGGAAGTTTGGATTCTCAGGGTTTGGGAATCCAAAT	1636
457	Db	CACCTTAGAGATGCTACAAAGTCCAGGAAGTTTGGATTCTCAGGGTTTGGGAATCCAAAT	398
1637	Qy	GCNAATCTTAGTGAACACAGCAGCAACATCTGTGACAAATGTTTTCTTCCTCCAGGACCAAGGA	1696
397	Db	GCNAATCTTAGTGAACACAGCAGCAACATCTGTGACAAATGTTTTCTTCCTCCAGGACCAAGGA	338
1697	Qy	GGAGCCTTAGAGATGACATGTTTCCTTGAGTCTCTACTATGCTGAAGGAA--TTTTTTTTTT	1754
337	Db	GGAGCCTTAGAGATGACATGTTTCCTTGAGTCTCTACTATGCTGAAGGAAATTTTTTTTTTT	278
1755	Qy	CCCTTTTGGTAGCTATGGAAGGTTGGAAAATTTTGGAAAGTGGGGGACTAGGAATTTATGG	1814
277	Db	CCCTTTTGGTAGCTATGGAAGGTTGGAAAATTTTGGAAAGTGGGGGACTAGGAATTTATGG	218
1815	Qy	ACAAATAAGGTTCCATTCGATTTATTTCGATTTTTTGGTTTGGTTTGTGTGTTGTAATTTTATA	1874
217	Db	ACAAATAAGGTTCCATTCGATTTATTTCGATTTTTTGGTTTGGTTTGTGTGTTGTAATTTTATA	158
1875	Qy	CAGCCACAGGATGGTATAGTATATATCTAGTATTTCAAAA	1914
157	Db	CAGCCACAGGATGGTATAGTATATATCTAGTATTTTCAAA	118

RESULT 4	746 bp	linear	EST 24-MAY-2001
BE658511/c		mrna	
LOCUS			
GM700006A20H3		Glycine max	
DEFINITION		clone Gm-r1070-2262 3',	
		mrna sequence.	

ACCESSION	BB658511
VERSION	BB658511.1
KEYWORDS	GI:9984403
SOURCE	E8T.
ORGANISM	Glycine max (soybean)
	Glycine max
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.

REFERENCE	AUTHORS	TITLE	JOURNAL	COMMENT
1	(bases 1 to 746) Vodkin, L., Keim, P., Shoemaker, R., Retzel, E., Khanna, A., Corvelli, V., Expanding, J., Raph, C., Shoop, E., Pardinias, J., Liu, L. and Lewin, H.	A Functional Genomics Program for Soybean (NSF 9872565)	Unpublished	
	Other_ESTRs: AW133440 corresponding to Gm-cl015-1643 (5') Contact: Vodkin, L.O., PI, A Functional Genomics Program for Soybean (NSF 9872565)			
	Lewin, H. A., Director, Keck Center for Comparative and Functional Genomics University of Illinois			

Edwin R. Madigan Building, 1201 W. Gregory, Urbana, IL 61801, USA
Tel: (217) 244-6147
Fax: (217) 333-4582
Email: 1-vodkin@uiuc.edu
This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134. For further information
call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314)
427-3324 or contact: clones@genomesystems.com or info@genome
systems.com web site: www.genomesystems.com
Seq primer: 5'-~~TT~~(A/C/G)-3'.

FEATURES

```

1. .746
source
/organism="Glycine max"
/mol_type="mRNA"
/db_xref="taxon:3847"
/clone="Gm-r1070-2262"
/clone_lib="Gm-r1070"
/note="The library Gm-r1070 is a sequence-driven, reracked
set of 9,216 clones selected from cDNA libraries from that
various tissues and stages of development of soybean that
represent 2,639 sequences from immature cotyledons, 1,770
from immature seed coats, 3,938 from flowers, and 869
from young pods. The 5' ESTs of the source clones form
the different libraries was used to select singletons, or
a representative of each contig, which were reracked to
form library Gm-r1070. The cDNA clones of the reracked
Gm-r1070 library were then sequenced at the 3' end. The
contig analysis to select unique genes was performed by
the laboratory of Ernest Ritzel, Center for Computational
Genomics and Bioinformatics, University of Minnesota,
http://www.cbc.umn.edu/ResearchProjects/Soybean/index.html
. Reracking was performed by Genome Systems, St. Louis,
http://www.genomesystems.com, and 3' sequencing by the
Keck Center for Comparative and Functional Genomics,
University of Illinois,
http://www.life.uiuc.edu/biotech/keck.html. Note: The
corresponding 5' EST from each clone in the Gm-r1070
library is listed in the 'OTHER EST' field. The detailed
information on the source library for each clone can also
be obtained by referring to the Genome Systems clone ID of
the original cDNA library that is also listed under
'OTHER EST'."

```

BASE COUNT	216 a	151 c	154 g	198 t	27 others
ORIGIN					

Query Match	30.3%;	Score 583.8;	DB 10;	Length 746;
Best Local Similarity	95.1%;	Pred. No. 8.5e-93;		
Matches 588;	Conservative 0;	Mismatches 30;	Indels 0;	Gaps 0;
Qy	1297	ACAAGTGCACACACCCAGGATGTC	CGATGAGGAAAGCACGCTGGAAAGAGCCTCATGACCC	1356
Db	746	ACNAGTGCACANNCCAGGATNTTCC	AGTGANNAGCACGTTNNAAGANNNNCACANNANN	687
Qy	1357	TAAGGGCTGTATCACAACCTATAG	GGGAAGCACACCAATCATGATGTTCTTCGACGCCGTG	1416
Db	686	NAAGGGCTGTGATCACAACTTAT	GAGGGAAGCACCAACCATGATGTTNNNGACGCCGTG	627
Qy	1417	GCAGTGCAGCCATTTCTGTGAA	CAGACCAATGCCTTCAAAACCCACACCAACA	1476
Db	626	GCAGTGCAGCCATTCTGTAA	CAGACCAATGCCAAACATGCTTCAAAACCCACACCAACA	567
Qy	1477	CTGCAGCCATTTCGGTAAGGCTCT	TTCGACGTGATCCACAAAGATGACAAATTCCTTCAGA	1536
Db	566	CTGCAGCCATTTCGGTAAGGCTCT	TTCGACGTGATCCACAAAGTGCAAATTCCTTCAGA	507
Qy	1537	ACCAAGATCACAAAGCACACCAAG	AGGGCAATCACCCCTTAGAGATGCTACAAA	1596
Db	506	ACCAAGATCACAAAGCACACCAAG	AGGGCAATCACCCCTTAGAGATGCTACAAA	447
Qy	1597	GTCAGGAAGTTTGGATTCTCAGG	GTTCGGGAATCCAAATGCAATCTTACGTGAACGAC	1656
Db	446	GTCAGGAAGTTTGGATTCTCAGG	GTTCGGGAATCCAAATGCAATCTTACGTGAACGAC	387

RESULT 6
 BQ295846
 LOCUS
 DEFINITION
 BQ295846 603 bp mRNA linear EST 16-MAY-2002
 sso29f11.y1 Gm-cl054 Glycine max cDNA clone SOYBEAN CLONE ID:
 Gm-cl054-5637 5' similar to TR:Q40090 Q40090 SPFI PROTEIN. ;, mRNA
 sequence.
 ACCESSION
 VERSION BQ295846.1 GI:20811368
 KEYWORDS
 SOURCE EST.
 ORGANISM
 Glycine max (soybean)
 Glycine max
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 ; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Glycine.
 1 (bases 1 to 603)
 Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
 A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
 Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
 Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
 R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
 R., Waterston,R. and Wilson,R.
 Public Soybean EST Project
 Unpublished
 Contact: Shoemaker R/Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watson.wustl.edu
 This clone is available through: ResGen, Invitrogen Corp. 2130
 South Memorial Parkway Huntville, AL 35801 For further information
 call: (800)-533-4363 or contact: ccu@resgen.com web site:
 www.resgen.com
 Seg primer: -4ORF from Gibco
 High quality sequence stop: 421.
 Location/Qualifiers
 1..603
 /organism="Glycine max"
 /mol_type="mRNA"
 /db_xref="taxon:3847"
 /clone="SOYBEAN CLONE ID: Gm-cl054-5637"
 /tissue_type="Leaf, 3 week old, greenhouse grown"
 /lab_host="DH10B"
 /clone_lib="Gm-cl054"
 /note="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2:
 XhoI; The Harosoy NIL was constructed and seed was
 provided by Dr. J. Specht, University of Nebraska
 (Shoemaker and Specht, 1995). The cDNA library was
 constructed from mRNA isolated from leaf tissue at various
 developmental stages of 3 week old greenhouse grown
 plants. Complementary DNA was synthesized from mRNA using
 a primer consisting of a poly(dT) sequence with a XhoI
 restriction site and a 3' anchor. EcoRI adapters were
 ligated to the blunt-ended cDNA fragments followed by
 XhoI digestion. The cDNA fragments were directionally
 cloned into the EcoRI-XhoI restriction site of the
 pBluescript vector. The ligated cDNA fragments were
 transformed into DH10B host cells (GibcoBRL). This
 library was constructed in cooperation with Dr. Paul
 Kelm's laboratory at Northern Arizona University."
 217 a 116 c 136 g 134 t
 BASE COUNT
 ORIGIN
 Query Match 30.0%; Score 579; DB 13; Length 603;
 Best Local Similarity 97.5%; Pred. No. 6.5e-92;
 Matches 588; Conservative 0; Mismatches 15; Indels 0; Gaps 0;
 755 GGGAGTGAAGTCCAAAGTATTATTAACAATGCACATACCCCAATTCCTTCAAGAA 814
 |||||

Db 1 GGGAGTGAAGTCCAAAGTATTATTAACAATGCACATACCCCAATTCCTTCAAGAA 60
 815 GAAGGTTGAGAGTCTTTAGATGGACAAATTAAGTGGATAGTATTTATAAGGGTACTCATAA 874
 |||||
 Db 61 GAAGGTTGAGAGTCTTTAGATGGACAAATTAAGTGGATAGTATTTATAAGGGTACTCATAA 120
 |||||
 Qy 875 CCATCCTAAGCCTCAAAATATCTAGGAGAAACTCATCAAACTCCTCTTCTTCTTGGCAATCCC 934
 |||||
 Db 121 CCATCCTAAGCCTCAAAATATCTAGGAGAAACTCATCAAACTCCTCTTCTTCTTGGCAATCCC 180
 |||||
 Qy 935 TCATTCAAAATTCATCAGAACTGAAATCCAGATCAATCCTATGCCACACATGGAAGTGG 994
 |||||
 Db 181 TCATTCAAAATTCATCAGCGCTGAAATCCAGATCAATCCTATGCCACACATGGAAGTGG 240
 |||||
 Qy 995 ACAATATGGATTTCAGCTGCCACCCAGAAACTCATCAATATCAATTTGGAGATGATGATTT 1054
 |||||
 Db 241 ACAATATGGATTTCAGCTGCCACCCAGAAACTCATCAATATCAATTTGGAGATGATGATTT 300
 |||||
 Qy 1055 TGAGCAGAGTTCCTCAAAAGTAAATCAGAGGGGATGAATATGATGAAGATGAACCTGA 1114
 |||||
 Db 301 TGAGCAGAGTTCCTCAAAAGTAAATCAGAGGGGATGAATATGATGAAGATGAACCTGA 360
 |||||
 Qy 1115 TGCCAAAGATGGAATAATGAAGTGAAATGAGGTGTAATGAGGTGATGAGCCCTGGAGTAGAAC 1174
 |||||
 Db 361 TGCCAAAGATGGAATAATGAAGTGAAATGAGGTGTAATGAGGTGATGAGCCCTGGAGTAGAAC 420
 |||||
 Qy 1175 AGTGAGAGAACCTTAGAGTTGTAGTTTCAGACCAACAGTGCATGATGATATCTTGTATGATGG 1234
 |||||
 Db 421 AGTAAGAGAACCTTAGAGTTGTAGTTTCAGACCAACAGTGCATGATGATATCTTGTATGATGG 480
 |||||
 Qy 1235 CTATAGGTGGAGAAATACGGGCAGAAAGTAGTGAAGGCAATCCAAATCCAAAGGAGTTA 1294
 |||||
 Db 481 GTATAGATGAGAAATACGGGCAGAAAGTAGTGAAGGCAATCCAAATCCAAAGGAGTTA 540
 |||||
 Qy 1295 CTCAAGTGCACACACCCAGGATGTCAGTGAGGAGCAGCTGGAAGGCCCTCAGATGA 1354
 |||||
 Db 541 CTCAAGTGTACACACCCAGGATGTCAGTGAGGAGCAGCTGGAAGGCCCTCAGATGA 600
 |||||
 Qy 1355 CCT 1357
 |||||
 Db 601 CCT 603
 |||||
 RESULT 7
 BQ080806 572 bp mRNA linear EST 04-APR-2002
 banl0e08.y1 Gm-cl084 Glycine max cDNA clone SOYBEAN CLONE ID:
 Gm-cl084-3663 5' similar to TR:Q40090 Q40090 SPFI PROTEIN. ;, mRNA
 sequence.
 ACCESSION BQ080806
 VERSION BQ080806.1 GI:19936302
 KEYWORDS EST.
 SOURCE Glycine max (soybean)
 ORGANISM Glycine max
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
 ; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
 Glycine.
 1 (bases 1 to 572)
 Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
 A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
 Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
 Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
 R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
 R., Waterston,R. and Wilson,R.
 Public Soybean EST Project
 Unpublished
 Contact: Shoemaker R/Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 TITLE
 JOURNAL
 COMMENT

Email: est@watson.wustl.edu
This clone is available through: ResGen, Invitrogen Corp. 2130
South Memorial Parkway Huntsville, AL 35801 For further information
call: (800)-533-4363 or contact: ccu@resgen.com web site:
www.resgen.com

Seq primer: -40RP from Gibco
High quality sequence stop: 421.

FEATURES

Location/Qualifiers
1..572
/organism="Glycine max"
/mol_type="mRNA"
/db_xref="taxon:3847"
/clone="SOYBEAN CLONE ID: Gm-c1084-3663"
/tissue_type="Etiolated hypocotyls (Williams 82)"
/lab_host="DH10B"
/clone_lib="Gm-c1084"
/note="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2: XhoI; The cDNA library was constructed by M. Bhattacharya from mRNA isolated from etiolated hypocotyls from the cultivar Williams 82. Tissue was inoculated with Phytophthora sojae race 1 and tissues were harvested 2 and 4 hours following infection. The library is the pool of these two time points. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with a XhoI restriction site. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into DH10B host cells (Gibco BRL). This library was constructed by M. Bhattacharya in the laboratory of Dr. Randy Shoemaker at Iowa State University."
204 a 112 c 120 g 136 t

BASE COUNT 204 a 112 c 120 g 136 t

ORIGIN

Query Match 29.5%; Score 568.8; DB 13; Length 572;
Best Local Similarity 99.7%; Pred. No. 4.1e-90;
Matches 570; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 607 GTTTTTCCTGAGATTGCTGCTCAAACTAACCATAGCATGGTTTCATCCGATT 666
DB 1 GTTTTTCCTGAGATTGCTGCTCAAACTAACCATAGCATGGTTTCATCCGATT 60

QY 667 ATGCGAATTAACCCCAATCTCAGACTTTAAGTAGAAGGTGAGTGGTACAATT 726
DB 61 ATGCGAATTAACCCCAATCTCAGACTTTAAGTAGAAGGTGAGTGGTACAATT 120

QY 727 GAGGAAATATGCCAAAACCAAGTGAAGGAAAGTGAATTCAGAGTTATTACAAT 786
DB 121 GAGGAAATATGTTCAAAAACCAAGTGAAGGAAAGTGAATTCAGAGTTATTACAAT 180

QY 787 GCACATACCCCAATTCGCTTACAAAGAAAGGTTGAGAGTCTTTAGATGGCAATTA 846
DB 181 GCACATACCCCAATTCGCTTACAAAGAAAGGTTGAGAGTCTTTAGATGGCAATTA 240

QY 847 CTGAGATAGTTTATAAGGGTACTCATTAACCATCTTAAGCCTCAAAATCTAGGAGAACT 906
DB 241 CTGAGATAGTTTATAAGGGTACTCATTAACCATCTTAAGCCTCAAAATCTAGGAGAACT 300

QY 907 CATCAAACTCCTCTTCTTTGCAATCCCTCATTTCAAAATTCATCAGAACTGAAATCCCAG 966
DB 301 CATCAAACTCCTCTTCTTTGCAATCCCTCATTTCAAAATTCATCAGAACTGAAATCCCAG 360

QY 967 ATCAATCCTATGCCACATGGAGTGGACAAATGGATTGAGTCCACCCAGAAACT 1026
DB 361 ATCAATCCTATGCCACATGGAGTGGACAAATGGATTGAGTCCACCCAGAAACT 420

QY 1027 CATCAATATCAATTGGAGATGATGATTTTGAGCAGAGTTCCTCAAAAGTGTAAATCAGGAG 1086
DB 421 CATCAATATCAATTGGAGATGATGATTTTGAGCAGAGTTCCTCAAAAGTGTAAATCAGGAG 480

QY 1087 GGGATGAATATGATGAAGATGAACCTGATGCCAAAAGATGAAAATTTGAAGGTGAAAATG 1146

Db 481 GGGATGAATATGATGAAGATGAACCTGATGCCAAAAGATGAAAATTTGAAGGTGAAAATG 540

QY 1147 AGGGTATGTCAGCCCTGAGAGTAGAAGTGAACAGTG 1178
|||||
Db 541 AGGGTATGTCAGCCCTGAGAGTAGAAGTGAACAGTG 572
|||||

RESULT 8
BU091659 568 bp mRNA linear EST 29-AUG-2002
LOCUS ST7506.y1 Gm-c1054 Glycine max cDNA clone GENOME SYSTEMS CLONE ID:
DEFINITION Gm-c1054-323 5' similar to TR:Q40090 Q40090 SPFI PROTEIN. ; mRNA
sequence.
ACCESSION BU091659
VERSION BU091659
KEYWORDS EST.
SOURCE Glycine max (soybean)
ORGANISM Glycine max

REFERENCE
1 (bases 1 to 568)
Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Wilson,R.
Public Soybean EST Project
Unpublished
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: Genome Systems, Inc. 4633 World
Parkway Circle St. Louis, Missouri 63134 For further information
call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or
info@genomesystems.com web site: www.genomesystems.com
High quality sequence stop: 442.

FEATURES
Location/Qualifiers
1..568
/organism="Glycine max"
/mol_type="mRNA"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-c1054-323"
/tissue_type="Leaf, 3 week old, greenhouse grown"
/lab_host="DH10B"
/clone_lib="Gm-c1054"
/note="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2: XhoI; The Harosoy NIL was constructed and seed was provided by Dr. J. Specht, University of Nebraska (Shoemaker and Specht, 1995). The cDNA library was constructed from mRNA isolated from leaf tissue at various developmental stages of 3 week old greenhouse grown plants. Complementary DNA was synthesized from mRNA using a primer consisting of a poly(dT) sequence with a XhoI restriction site and a 3' anchor. EcoRI adapters were ligated to the blunt-ended cDNA fragments followed by XhoI digestion. The cDNA fragments were directionally cloned into the EcoRI-XhoI restriction site of the pBluescript vector. The ligated cDNA fragments were transformed into DH10B host cells (GibcoBRL). This library was constructed in cooperation with Dr. Paul Keim's laboratory at Northern Arizona University."
198 a 121 c 130 g 119 t

BASE COUNT 198 a 121 c 130 g 119 t

Query Match 28.7%; Score 553.6; DB 13; Length 568;

Best Local Similarity 98.4%; Pred. No. 2e-87; Matches 559; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

Qy 899 GAGAACTCATCAAACTCTCTCTCTGATCCCTCATCAAAATCCATCAGAACTGA 958
Db 1 GAGAACTCATCAAACTCTCTCTCTGATCCCTCATCAAAATCCATCAGAACTGA 60
Qy 959 AATCCAGATCAATCTATGATCCACATGGAAGTGGCAAAATGATTCAGTGCACCCC 1018
Db 61 AATCCAGATCAATCTATGATCCACATGGAAGTGGCAAAATGATTCAGTGCACCCC 120
Qy 1019 AGAAATCTCATCAATATCAATGAGATGATGATTTTGGAGAGAGTTCCCAAAAGTGA 1078
Db 121 AGAAATCTCATATATCAATGAGATGATGATTTTGGAGAGAGTTCCCAAAAGTGA 180
Qy 1079 ATCAGGAGGATGAATATGATGAAGATGAACCTGATGCCAAAGATGGAAAATGAAG 1138
Db 181 ATCAGGAGGATGAATATGATGAAGATGAACCTGATGCCAAAGATGGAAAATGAAG 240
Qy 1139 TGAAATCAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1198
Db 241 TAAATATGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
Qy 1199 TCAGAACCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1258
Db 301 TCAGAACCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
Qy 1259 GAAATGATGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1318
Db 361 GAAATGATGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 420
Qy 1319 TCCAGTGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1378
Db 421 TCCAGTGGAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
Qy 1379 TGAGGAAAGCAACCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1438
Db 481 TGAGGAAAGCAACCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540
Qy 1439 CAGACCAATGCCAAACATGCTTCAAC 1466
Db 541 CAGACCAATGCCAAACATGCTTCAAC 568

RESULT 9
BI975061
LOCUS
DEFINITION
531 bp mRNA linear EST 30-NOV-2001
ID: Gm-cl068-4552 5' similar to TR:Q9SXP5 Q9SXP5 TRANSCRIPTION
FACTOR NTRK2. ; mRNA sequence.

ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
EST.
Glycine max (soybean)
Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
1 (bases 1 to 531)

REFERENCE
AUTHORS
Shoemaker, R., Kelm, P., Vodkin, L., Erpelting, J., Coryell, V., Khanna
A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,
Wylie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers
Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk
R., Rutter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann
R., Waterston, R., and Wilson, R.
Public Soybean EST Project
Unpublished

TITLE
JOURNAL
COMMENT
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800

Fax: 314 286 1810
Email: est@wustl.edu
This clone is available through: ResGen, Invitrogen Corp. 2130
South Memorial Parkway Huntville, AL 35801 For further information
call: (800)-533-4363 or contact via email: ccu@resgen.com
High quality sequence stop: 438.

FEATURES

Location/Qualifiers
source
1..531
/organism="Glycine max"
/mol_type="mRNA"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl068-4552"
/issue_type="Leaf, drought stressed, 1 month old plants,
greenhouse grown"
/lab_host="DH10B"
/clone_lib="Gm-cl068"
/notes="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2:
XhoI; The cDNA library was constructed from mRNA isolated
from drought stressed leaf tissue of the cultivar Williams
82. The month old greenhouse grown plants were deprived of
water for 3 days prior to harvesting the stressed leaf
tissue. Complementary DNA was synthesized from mRNA using
a primer consisting of a poly(dT) sequence with a XhoI
restriction site. EcoRI adapters were ligated to the
blunt-ended cDNA fragments followed by XhoI digestion. The
cDNA fragments were directionally cloned into the
EcoRI-XhoI restriction site of the pBluescript vector. The
ligated cDNA fragments were transformed into DH10B host
cells (GibcoBRL). This library was constructed in the
laboratory of Dr. Randy Shoemaker."
BASE COUNT 163 a 134 c 117 g 117 t
ORIGIN

Query Match 27.4%; Score 528.4; DB 12; Length 531;
Best Local Similarity 99.8%; Pred. No. 5.4e-83;
Matches 529; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1246 GAAATATCGGCGAGAAAGTAGTGAAGGGCAATCCAAATCCAAAGGAGTTACTACAAGTGA 1305
Db 2 GAAATATCGGCGAGAAAGTAGTGAAGGGCAATCCAAATCCAAAGGAGTTACTACAAGTGA 61
Qy 1306 CACACCCAGATGTCAGTGAAGAGACGTGGAAGAGCCTCAGTACCTAAGGGCTG 1365
Db 62 CACACCCAGATGTCAGTGAAGAGACGTGGAAGAGCCTCAGTACCTAAGGGCTG 121
Qy 1366 TGATCACTTATGAGGAAAGCAGACCATCATGTTCTCGAGCCGCGGAGTGGCA 1425
Db 122 TGATCACTTATGAGGAAAGCAGACCATCATGTTCTCGAGCCGCGGAGTGGCA 181
Qy 1426 GCCATTCTGTGAACAGACCAATGCCAAACAAATGCTTCAAAACCAACACCACTGCAGCCA 1485
Db 182 GCCATTCTGTGAACAGACCAATGCCAAACAAATGCTTCAAAACCAACACCACTGCAGCCA 241
Qy 1486 CTTCCGTAAGGCTCTTGCAGTGCATCCACAAAGTGAACAATCCCTTCAAGAACCAAGAT 1545
Db 242 CTTCCGTAAGGCTCTTGCAGTGCATCCACAAAGTGAACAATCCCTTCAAGAACCAAGAT 301
Qy 1546 CACAGCACCAACAGAGGCGCATCCCTTCAACCTAGAGATGCTACAAGTCCAGGAA 1605
Db 302 CACAGCACCAACAGAGGCGCATCCCTTCAACCTAGAGATGCTACAAGTCCAGGAA 361
Qy 1606 GTTTTGGATTCTCAGGGTTTGGGAATCCAAATGCAATCTTACGTGAACAGCAGCAATAT 1665
Db 362 GTTTTGGATTCTCAGGGTTTGGGAATCCAAATGCAATCTTACGTGAACAGCAGCAATAT 421
Qy 1666 CTGCAATGTTTCTCTCTCCAGGACCAAGAGGAGGCTAGAGATGATGATGATGATGAT 1725
Db 422 CTGCAATGTTTCTCTCTCCAGGACCAAGAGGAGGCTAGAGATGATGATGATGATGAT 481
Qy 1726 CTCTACTATGCTGAAGGAATTTTTTTTTTCCCTTTTGGTAGCTATGGAA 1775
Db 482 CTCTACTATGCTGAAGGAATTTTTTTTTTCCCTTTTGGTAGCTATGGAA 531


```
RESULT 10
BM519918      585 bp      mRNA      linear      EST 15-FEB-2002
LOCUS      sak86c03.y1 Gm-cl057 Glycine max cDNA clone SOYBEAN CLONE ID:
DEFINITION      Gm-cl057-2430 5' similar to TR:Q40090 Q40090 SPFI PROTEIN. ; mRNA
sequence.
ACCESSION      BM519918
VERSION      BM519918.1 GI:18691070
KEYWORDS      EST.
SOURCE      Glycine max (soybean)
ORGANISM      Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
1 (bases 1 to 585)
Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V., Khanna
,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C.,
Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers
,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk
,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann
,R., Waterston,R. and Wilson,R.
Public Soybean EST Project
Unpublished
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: ResGen, Invitrogen Corp. 2130
South Memorial Parkway Huntville, AL 35801 For further information
call: (800)-533-4363 or contact: cu@resgen.com web site:
www.resgen.com
Seq primer: -40RP from Gibco
High quality sequence stop: 425.
Location/Qualifiers
1. 585
/organism="Glycine max"
/mol_type="mRNA"
/db_xref="taxon:3847"
/clone="SOYBEAN CLONE ID: Gm-cl057-2430"
/tissue_type="Degenerating cotyledons, 2 week old
seedling"
/lab_host="DH10B"
/clone_lib="Gm-cl057"
/notes="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2:
XhoI; The cDNA library was constructed from mRNA isolated
from degenerating cotyledons of 2 week old seedlings from
PI468916. Complementary DNA was synthesized from mRNA
using a primer consisting of a poly(dT) sequence with a
XhoI restriction site. EcoRI adapters were ligated to the
blunt-ended cDNA fragments followed by XhoI digestion.
The cDNA fragments were directionally cloned into the
EcoRI-XhoI restriction site of the pBluescript vector. The
ligated cDNA fragments were transformed into DH10B host
cells (GibcoBRL). This library was constructed in the
laboratory of Dr. Randy Shoemaker."
BASE COUNT      189 a 137 c 130 g 129 t
ORIGIN
Query Match      26.7%; Score 515; DB 12; Length 585;
Best Local Similarity 93.8%; Pred. No. 1.2e-80;
Matches 549; Conservative 0; Mismatches 30; Indels 6; Gaps 1;
Qy 1173 ACAGTGAGAGAACCTAGAGTTGTAGTTTCAGACACACAGTGCATTTGATATCTTGTATGAT 1232
Db 1 ACAGTAGAGAACCTAGAGTTGTAGTTTCAGACACACAGTGCATTTGATATCTTGTATGAT 60
Qy 1233 GGCTATAGTGGAGAAATACGGGCAGAAAGTAGTGAAGGCGCAATCCAAATCCAAAGGAGT 1292
Db 1173 GGCTATAGTGGAGAAATACGGGCAGAAAGTAGTGAAGGCGCAATCCAAATCCAAAGGAGT 1292
```

```
source
1. .507
/organism="Glycine max"
/mol_type="mRNA"
/db_xref="taxon:3847"
/clone="SOYBEAN CLONE ID: Gm-cl072-2949"
/tissue_type="seedlings induced for symptoms of SDS
(Sudden Death Syndrome) disease"
/dev_stage="2-3 weeks old"
/lab_host="DH10B"
/clone_lib="Gm-cl072"
/note="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2:
XhoI; The cDNA library was constructed from mRNA isolated
from 2-3 week old seedlings that were induced for symptoms
of SDS (Sudden Death Syndrome) disease by the
translocation of culture filtrate of Fusarium solani f.
sp. glycines (Plant Cell Report 18:375-380). Cultivar PI
567374 is partially resistant to the disease SDS. Plant
tissue (expanded leaves, folded leaves, and new shoots)
were collected at 1, 6, 24, and 48 hrs. after inoculation
and their mRNA pooled equally for cDNA construction. The
library was prepared using the Stratagene pBluescript II
SK(+) library construction kit. Complementary DNA was
synthesized from mRNA using a primer consisting of a
poly(dT) sequence with an XhoI restriction site. EcoRI
adaptors were ligated to the blunt-ended cDNA fragments
followed by XhoI digestion. The cDNA insert is protected
from XhoI digestion via methylation during first strand
synthesis. The cDNA fragments were directionally cloned
into the EcoRI-XhoI restriction site of the pBluescript
vector. The ligated cDNA fragments were transformed into
E. coli ElectroMax DH10B host cells. Plants were inoculated
by Shuxian Li (Glen Hartman lab, University of Illinois).
Library was constructed by Steve Clough (Lila Vodkin lab,
University of Illinois)."
BASE COUNT 169 a 118 c 96 g 124 t
ORIGIN
Query Match 26.2%; Score 505.4; DB 12; Length 507;
Best Local Similarity 99.8%; Pred. No. 6.2e-79; Indels 0; Gaps 0;
Matches 506; Conservative 0; Mismatches 1;

QY 306 CTCCTTAACTCTTCCAACTTCTGCCATCTCCAACTGAGGAGGATTTGTCCTCAGAGC 365
DB 1 CTCCTTAACTCTTCCAACTTCTGCCATCTCCAACTGAGGAGGATTTGTCCTCAGAGC 60
QY 366 TTCAATTGGAAGAGCAGTTTCAGGGGGGAATCAGCAAAATTTGCAAGGAGGAGCAAAAGC 425
DB 61 TTCAATTGGAAGAGCAGTTTCAGGGGGGAATCAGCAAAATTTGCAAGGAGGAGCAAAAGC 120
QY 426 TTCTCAAAATTTCTTTTCCAAACCCGATCAGAGCTCTGCTTCATCCACAGCAACATAC 485
DB 121 TTCTCAAAATTTCTTTTCCAAACCCGATCAGAGCTCTGCTTCATCCACAGCAACATAC 180
QY 486 CAGTCTTCAAAATGTCACAGTTTCAACACACAGCAGCATGGAGTTTTCAGGAGGCCACGAAA 545
DB 181 CAGTCTTCAAAATGTCACAGTTTCAACACACAGCAGCATGGAGTTTTCAGGAGGCCACGAAA 240
QY 546 CAGGATAATTTTCTCCTCAGGAAAGGGTATGATGAAAACTGAAAACTCTTCTTCATCGAG 605
DB 241 CAGGATAATTTTCTCCTCAGGAAAGGGTATGATGAAAACTGAAAACTCTTCTTCATCGAG 300
QY 606 AGTTTTTCCCTGAGATGCTAGTGTCCAACTAACCTACCTAGCATGGTTTCATTCGGAT 665
DB 301 AGTTTTTCCCTGAGATGCTAGTGTCCAACTAACCTACCTAGCATGGTTTCATTCGGAT 360
QY 666 TATGCAATTACCCCCCAATCTCAGACTTTTAAGTAGAAGGTGAGATGATGGGTACAAAT 725
DB 361 TATGCAATTACCCCCCAATCTCAGACTTTTAAGTAGAAGGTGAGATGATGGGTACAAAT 420
QY 726 TGGAGGAAATATGGCCAAAAACAAGTGAAGGGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 785
DB 421 TGGAGGAAATATGGCCAAAAACAAGTGAAGGGAAGTGAAGTGAAGTGAAGTGAAGTGAAGT 480
```

```
786 TGACATACCCCAATTGCTCTACAAAG 812
|||||
481 TGACATACCCCAATTGCTCTACAAAG 507

RESULT 12
BUS78260
LOCUS
DEFINITION
Gm-cl074-6003 5' similar to TR:Q40090 Q40090 SP1 PROTEIN. ; mRNA
sequence.
BUS78260
VERSION
BUS78260.1 GI:23057720
EST.
Glycine max (soybean)
Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids
; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
1 (bases 1 to 509)
Shoemaker, R., Keim, P., Vodkin, L., Erpelding, J., Coryell, V., Khanna
, A., Bolla, B., Marra, M., Hillier, L., Kucaba, T., Martin, J., Beck, C.,
Wyllie, T., Underwood, K., Steptoe, M., Theising, B., Allen, M., Bowers
, Y., Person, B., Swaller, T., Gibbons, M., Pape, D., Harvey, N., Schurk
, R., Ritter, E., Kohn, S., Shin, T., Jackson, Y., Cardenas, M., McCann
, R., Waterston, R. and Wilson, R.
Public Soybean EST Project
Unpublished
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: ResGen, Invitrogen Corp. 2130
South Memorial Parkway Huntville, AL 35801 For further information
call: (800)-533-4363 or contact: cu@resgen.com web site:
www.resgen.com
Seq primer: -40RP from Gibco
High quality sequence stop: 405.
Location/Qualifiers
1. .509
/organism="Glycine max"
/mol_type="mRNA"
/db_xref="taxon:3847"
/clone="SOYBEAN CLONE ID: Gm-cl074-6003"
/tissue_type="seedlings induced for HR (hypersensitive
response)"
/dev_stage="9-11 day old"
/lab_host="DH10B"
/clone_lib="Gm-cl074"
/note="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2:
XhoI; The cDNA library was constructed from mRNA isolated
from 9-11 day old seedlings that were induced for HR
(hypersensitive response) by vacuum infiltrating plant
tissue with Pseudomonas syringae pv. glycinea carrying the
avrB gene (Genetics 141:1597-1604). Plant tissue (expanded
unifoliate leaves) was collected at 2, 4, 8, 12, 24, 36,
and 53 hrs after inoculation and their mRNA pooled equally
for cDNA construction. The library was prepared using the
Stratagene pBluescript II SK(+) library construction kit.
Complementary DNA was synthesized from mRNA using a primer
consisting of a poly(dT) sequence with an XhoI restriction
site. EcoRI adaptors were ligated to the blunt-ended cDNA
fragments followed by XhoI digestion. The cDNA insert is
protected from XhoI digestion via methylation during first
strand synthesis. The cDNA fragments were directionally
cloned into the EcoRI-XhoI restriction site of the
pBluescript vector. The ligated cDNA fragments were
transformed into E. coli ElectroMax DH10B host cells. Plant
care, inoculations, and library construction were
```

16 AGGGCAATCCAAATCCAAAGGAGTTACTACAAATGTACACACCCCTAACTGTCCAGTGAGG 675

QY 1329 AAGCAGCTGGAAGAGCCTCACATGACCTAAGGCTGTGATCAAACTTATGAGGAAAG 1388
 |||||
 Db 676 AAGCATGTTGAAGCTGCTTGCATGATTTAAGGCTGTGATCACTACTTATGAGGAAA 735
 |||||
 QY 1389 CACAACCATGATGTTCTCGAGCCCGTGGCAGTGGCA 1425
 |||||
 Db 736 CATAACCATGATGTCCTGCTGCTGCTGGAAGCGAA 772
 |||||

RESULT 14
 CD042161
 LOCUS
 DEFINITION psHB039x09f 301841 psHB: Infected hypocotyl soybean host. 48 hrs post infection Phytophthora sojae cDNA clone SHB039F09 5, mRNA sequence.
 CD042161
 EST.
 SOURCE Phytophthora sojae
 ORGANISM Phytophthora sojae
 Eukaryota; stramenopiles; Oomycetes; Pythiales; Pythiaceae; Phytophthora.
 1 (bases 1 to 570)
 Tyler,B.M., Judelson,H.S., Gijzen,M., Dean,R.A. and Waugh,M.E.
 USDA-IFAFS: Expression of Phytophthora sojae genes during infection and propagation
 Unpublished
 CONTACT: Tyler B
 Tyler lab
 VBI
 1980 Pratt Dr., Blacksburg, VA 24061, USA
 Tel: 540-231-7318
 Email: bmttyler@vt.edu
 PCR Primers
 FORWARD: BK reverse
 Plate: 039 row: F column: 09
 Seq primer: BK reverse
 High quality sequence stop: 570.
 Location/Qualifiers
 1..570
 /organism="Phytophthora sojae"
 /mol_type="mRNA"
 /db_xref="taxon:67593"
 /clone="SHB039F09"
 /tissue_type="infected host tissue"
 /cell_line="P6497"
 /dev_stage="48 hour post infection"
 /clone_lib="psHB: Infected hypocotyl soybean host. 48 hrs post infection"
 /note="vector: pBK-CMV; Site 1: EcoRI; Site 2: XhoI; USDA-IFAFS: Expression of Phytophthora sojae genes during infection and propagation."
 BASE COUNT 152 a 161 c 98 g 159 t
 ORIGIN

Query Match 25.1%; Score 484.4; DB 14; Length 570;
 Best Local Similarity 92.1%; Pred. No. 2.9e-75;
 Matches 525; Conservative 0; Mismatches 36; Indels 9; Gaps 1;

QY 107 CTCTGACCTCTGCTTCTCCCTTGGACAAACAAAGCCACACAG-----GGTGG 157
 |||||
 Db 1 CTCTGACCTTCTTCTCTCCACAGACACAAACCAACCACTGGTGGTGGTGG 60
 |||||

QY 158 TTCTCTGAGAGAACTGGCTGTGGTGTCCAAATCAAGTCCACACACACCTTCTCT 217
 |||||
 Db 61 TTCTCTGAGAGAACTGGCTGTGGTGTCCCAATCAAGTCCACACACCTTCTCT 120
 |||||

QY 218 GCCTCTCTCCCTCCCACTTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 277
 |||||
 Db 121 GCCTCTCTCTCCCTCCCACTTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 180
 |||||

QY 278 CCCTGCTGAGCTTCTTCTGACTCGCGGTCTCTCTTAACTCTTCCAACTTCTGCCATCTCC 337
 |||||

Db 181 CCCTGCTGAGCTTCTTGACTCCTCCTTAACTTCTCCACATTTCTGCCATCTCC 240
 |||||
 QY 338 AACAACCTGGAGCATTTGTTGCTCAGAGCTTCAATTGGAAGAGCAGTTTCAAGGGGGAATCA 397
 |||||
 Db 241 AACAACCTGGAGCATTTGTTGCTCGGAGCTTCAATTGGAAGAGCAGTTTCCGGGGGAACCA 300
 |||||
 QY 398 GCAAATTTGCAAGGAAGAGACAAAGCTTCTCAAAATTTCTCTTTCCAAACCCGATCAG 457
 |||||
 Db 301 GCGAATTTGCAAGGAAGAGACAAAGCTTCTCAAAATTTCTCTTTTCAAAACCCCAAGG 360
 |||||

QY 458 ACCTCTGCTTTCATCCACAGCAACATACCATGCTTCAAAATGTCACAGTTTCAACACAAACA 517
 |||||
 Db 361 ACCTCTGCTTTCATCCACAGCAACATATCATGCTTCTCAAAATGTCACAGTTTCAACACAAACA 420
 |||||

QY 518 GCCATGAGCTTTCAGGAGGCCACGAACACAGATATTTTCTCTCAGGAAGGATGAT 577
 |||||
 Db 421 GCCATGAGCTTTCAGGAGGCCACGAACAGCAGGATATTTTCTTTCAGGAAGAGCATGAT 480
 |||||

QY 578 GAAACTGAAACTCTCTTCCATGACAGATTTTTCCTCCCTGAGATTTGCTAGTGTCCAAAC 637
 |||||

Db 481 GAAACTGAAACTCTTTCATCCATGCAAGTTTTCCTCCCTGAGATTTGCTAGTGTCCAAAC 540
 |||||

QY 638 TAACCATAGCAATGGGTTTCAATCCGATTA 667
 |||||

Db 541 TAATCATAGCAATGGGTTTCAATCCGATTA 570
 |||||

RESULT 15
 BM187988
 LOCUS
 DEFINITION BM187988.1 GI:17518946
 saj83d08.y1 Gm-cl074 Glycine max cDNA clone SOYBEAN CLONE ID: Gm-cl074-3232 5' similar to TR:081639 O81639 ZINC FINGER PROTEIN.
 // mRNA sequence.
 BM187988
 EST.
 VERSION BM187988.1
 KEYWORDS Glycine max (soybean)
 SOURCE Glycine max
 ORGANISM Glycine max
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids ; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
 1 (bases 1 to 518)
 Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Coryell,V., Khanna A., Bolla,B., Maira,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers ,Y., Person,B., Swaller,T., Gibbons,M., Fape,D., Harvey,N., Schurr ,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann ,R., Waterston,R. and Willson,R.
 Public Soybean EST Project
 Unpublished
 TITLE Public Soybean EST Project
 JOURNAL
 COMMENT Contact: Shoemaker R/Public Soybean EST Project
 Public Soybean EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: estowatson.wustl.edu
 This clone is available through: ResGen, Invitrogen Corp. 2130 South Memorial Parkway Huntville, AL 35801 For further information call: (800)-533-4363 or contact: ccu@resgen.com web site: www.resgen.com
 Seq primer: -40RP from Gibco
 High quality sequence stop: 426.
 Location/Qualifiers
 1..518
 /organism="Glycine max"
 /mol_type="mRNA"
 /db_xref="taxon:3847"
 /clone="SOYBEAN CLONE ID: Gm-cl074-3232"
 /tissue_type="seedlings induced for HR (hypersensitive response)"

```
/dev_stage="9-11 day old"
/lab_host="DH10B"
/clone_lib="Gm-cl074"
/notes="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2:
XhoI; The cDNA library was constructed from mRNA isolated
from 9-11 day old seedlings that were induced for HR
(hypersensitive response) by vacuum infiltrating plant
tissue with Pseudomonas syringae pv. glycinea carrying the
avrB gene (Genetics 141:1597-1604). Plant tissue (expanded
unifoliate leaves) was collected at 2, 4, 8, 12, 24, 36,
and 53 hrs after inoculation and their mRNA pooled equally
for cDNA construction. The library was prepared using the
Stratagene pBluescript II SK(+) library construction kit.
Complementary DNA was synthesized from mRNA using a primer
consisting of a poly(dT) sequence with an XhoI restriction
site. EcoRI adaptors were ligated to the blunt-ended cDNA
fragments followed by XhoI digestion. The cDNA insert is
protected from XhoI digestion via methylation during first
strand synthesis. The cDNA fragments were directionally
cloned into the EcoRI-XhoI restriction site of the
pBluescript vector. The ligated cDNA fragments were
transformed into E. coli ElectroMax DH10B host cells. Plant
care, inoculations, and library construction were
performed by Steve Clough (Lila Vodkin lab, University of
Illinois)."
```

BASE COUNT 172 a 124 c 118 g 104 t
ORIGIN

```
Query Match 24.5%; Score 471.8; DB 12; Length 518;
Best Local Similarity 94.8%; Pred. No. 5e-73;
Matches 488; Conservative 0; Mismatches 27; Indels 0; Gaps 0;

Qy 1150 GTATGTCAGCCCTGGAAAGTAGAACAGTGAAGAACCTAGAGTTGTAGTTTCAGACCAACCA 1209
Db 1 GTATGTCGCCCTGGAAAGTAGAACAGTGAAGAACCTAGAGTTGTAGTTTCAGACCAACCA 60

Qy 1210 GTGACATTGATATCCTTTGATGATGGCTATAGTGGAGAAAATACGGGCAGAAAGTAGTGA 1269
Db 61 GTGACATTGATATCCTAGATGATGGGTATAGATGGAGAAAATACGGGCAGAAAGTAGTGA 120

Qy 1270 AGGCAATCCAAATCCAGAGTTACTACAGTGCACACCCAGGATGTCCAGTGAGGA 1329
Db 121 AGGCAATCCAAATCCAGAGTTACTACAGTGCACACCCAGGATGTCCAGTGAGGA 180

Qy 1330 AGCAGTGGAAAGAGCCTCATGACCTTAAGGGCTGTGATCAAACTTATGAGGGAAGC 1389
Db 181 AGCATGTGGAAAGAGCCTCATGACCTTAAGGGCTGTGATCAAACTTATGAGGGAAGC 240

Qy 1390 ACAACCATGATGTTCTTCGACCCCGTGGCAGTGGCAGCCATTCTGTGAACAGACCAATGC 1449
Db 241 ACAACCATGATGTTCTTCGACCTCGTGGCAGTGGCAGCCATTCTGTGAACAGACCAATGC 300

Qy 1450 CAACCAATGTTCAAAACACCAACCACTGACGCCACTTCGGTAAGSCTCTTCCAGTGA 1509
Db 301 CAACCAATGTTCAAAACACCAACCACTGACGCCACTTCGGTAAGSCTCTTCCAGTGA 360

Qy 1510 TCCACCAAGTGACAAATTCCTTCAGAAACCAAGATCAACAGCACCACCAAGAGGCAAT 1569
Db 361 TCCACCAAGTGACAAATTCCTTCAGAAACCAAGATCAACAGCACCACCAAGAGGCAAT 420

Qy 1570 CACCCCTTACCCTAGAGATGCTACAAAGTCCAGGAAGTTTGGATTCTCAGGGTTTGGGA 1629
Db 421 CACCCCTTACCCTAGAGATGCTACAAAGTCCAGGAAGTTTGGATTCTCAGGGTTTGGGA 480

Qy 1630 ATCCAATGCAATCTTACGTGAACAGCAGCACTA 1664
Db 481 ATCCAATGCAATCTTACATGAATCAGCAGCACTA 515
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 20, 2004, 16:16:52 ; Search time 47 Seconds

(without alignments)
1941.868 Million cell updates/sec

Title: US-09-890-811B-10

Perfect score: 3050

Sequence: 1 MASSSGSLDTASANSFTNF.....FSSRTKEPRDDMFLESLIC 575

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_19Jun03.*
1: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	3050	100.0	575	22 AAE05092	Soybean SPFI-relat
2	1605	52.6	560	23 ABG76909	Sunflower WRKY pol
3	1295	42.5	487	22 AAE05091	Rice SPFI-related
4	1139.5	37.4	512	21 AAG31843	Arabidopsis thalia
5	1136.5	37.3	503	21 AAG31844	Arabidopsis thalia
6	1117.5	36.6	488	22 AAE05090	Rice SPFI-related
7	904.5	29.7	619	22 AAE05093	Wheat SPFI-related
8	888.5	29.1	278	23 ABG76918	Wheat WRKY polypep
9	884	29.0	309	21 AAG46847	Arabidopsis thalia

10	861.5	28.2	514	24 ABP81179	Arabidopsis thalia
11	858	28.1	577	23 ABG76916	Soybean WRKY polyp
12	852	27.9	687	23 AAU92971	Arabidopsis transc
13	848.5	27.8	513	24 ABP81191	Arabidopsis thalia
14	818.5	26.8	279	21 AAG46848	Arabidopsis thalia
15	742	24.3	568	23 AAU92967	Arabidopsis transc
16	702.5	23.0	200	21 AAG42218	Arabidopsis thalia
17	702.5	23.0	216	21 AAG42217	Arabidopsis thalia
18	702.5	23.0	216	21 AAG46849	Arabidopsis thalia
19	625.5	20.5	406	23 ABG76910	Sunflower WRKY pol
20	619.5	20.3	413	23 ABG76911	Sunflower WRKY pol
21	618.5	20.3	583	23 ABG76922	Maize WRKY polypep
22	606	19.9	430	23 ABG76912	Sunflower WRKY pol
23	480	15.7	148	21 AAG42219	Arabidopsis thalia
24	460.5	15.1	258	23 ABG76920	Maize WRKY polypep
25	406.5	13.3	189	21 AAG23979	Arabidopsis thalia
26	400	13.1	163	21 AAG23980	Arabidopsis thalia
27	363	11.9	68	23 AAO18478	A thaliana WRKY33
28	353	11.6	150	22 AAE05089	Corn SPFI-related
29	353	11.6	173	23 ABG76915	Soybean WRKY polyp
30	319.5	10.5	170	22 AAE05088	Corn SPFI-related
31	318	10.4	191	21 AAG17496	Arabidopsis thalia
32	318	10.4	191	22 AAB67840	Amino acid sequenc
33	318	10.4	191	23 AAU92989	Arabidopsis transc
34	318	10.4	210	21 AAG17495	Arabidopsis thalia
35	316.5	10.4	528	23 AAU93139	Arabidopsis transc
36	315.5	10.3	318	22 AAE01958	Arabidopsis thalia
37	315.5	10.3	318	23 AAU93117	Arabidopsis transc
38	310.5	10.2	553	23 AAU93116	Arabidopsis transc
39	309.5	10.1	326	23 AAU92993	Arabidopsis transc
40	309.5	10.1	536	23 AAU92991	Arabidopsis transc
41	305.5	10.0	528	24 AAE30067	Arabidopsis thalia
42	301	9.9	185	24 ABP96034	Nicotiana tabacum
43	301	9.9	415	21 AAG45532	Arabidopsis thalia
44	301	9.9	545	21 AAG45531	Arabidopsis thalia
45	300.5	9.9	406	21 AAG45533	Arabidopsis thalia

ALIGNMENTS

RESULT 1
AAE05092
ID AAE05092 standard; Protein; 575 AA.
AC AAE05092;
XX
XX 12-SEP-2001 (first entry)
DT
DE Soybean SPFI-related transcription factor #5.
XX
XX Soybean; SPFI-related transcription factor; transgenic plant;
KW genetical mapping; physical mapping; plant breeding.
XX
XX Glycine max.
OS
XX
XX WC200149840-A2.
PN
XX
PD 12-JUL-2001.
XX
XX 22-DEC-2000; 2000WO-US35310.
XX
XX 04-JAN-2000; 2000US-0174325.
XX
XX (DUPO) DU PONT DE NEMOURS & CO E I.
PA
XX
XX Zhu Q, Famodu OO, Rafalski JA;
XX
XX WPI; 2001-441876/47.
DR
XX
DR N-PSDB; AAD09829.
XX
XX Novel isolated SPFI-related transcription factor polypeptides and
PT polynucleotides useful for producing transgenic plants -

Db	285	SSSSASNTLQMGAQSNHVDHDPQSVYHGSGQVDSVTPPNSNISVGDDSPDRS----	343
Qy	355	CKSGGD--EYDEPDPAKEWKIEGENGMS-APGSRTVREPRVVVQTSDIDLDGYRW	411
Db	342	-RSGDGVTVDEDEPEAKWKV-SENEGISMIGTKTVREPRIVVQTTSDIDLDGYRW	399
Qy	412	RKYGQKVKVGNPNPSYKYCTHPGCPVRKHVERASHDLRAVITYTEGKHNHDVPAARGSG	471
Db	400	RKYGQKVKVGNPNPSYKYCTSLGCSVRKHVERASQDLRSVITYTEGKHNHDVPMARGSG	459
Qy	472	SHSVNRPMPNNASNTHTAATSVRLLPVH-QSDNSL--QNORSQAPP--EGOSPFTLEM	526
Db	460	HRLQASTLSNAPSMT-----IKPMALSHYQVDSNSWVDTPRGPRYPSPSENQAPFTLEM	513
Qy	527	LQSPGSGFGSGFNPMQSVNQOQLSDNVFSRSTKEEPRDDMFLESLL	574
Db	514	LQSSDNFKYSRFENALKSNYNEHN-SERTFST-TKEEPRDDMFLESLL	559
RESULT 3			
AAE05091	ID	AAE05091 standard; Protein; 487 AA.	
XX	AC	AAE05091;	
XX	DT	12-SEP-2001 (first entry)	
XX	DE	Rice SPPI-related transcription factor #4.	
XX	KW	Rice; SPPI-related transcription factor; transgenic plant;	
KW	KW	genetical mapping; physical mapping; plant breeding.	
XX	OS	Oryza sativa.	
XX	BN	WO200149840-A2.	
XX	PD	12-JUL-2001.	
XX	PF	22-DEC-2000; 2000WO-US35310.	
XX	PR	04-JAN-2000; 2000US-0174325.	
PA	PA	(DUPO) DU PONT DE NEMOURS & CO E I.	
PI	PI	Zhu Q, Famodu OO, Rafalski JA;	
DR	DR	WPI; 2001-441876/47.	
DR	DR	N-PSDB; AAD09828.	
PT	PT	Novel isolated SPF-1 related transcription factor polypeptides and	
PT	PT	polynucleotides useful for producing transgenic plants -	
PS	PS	Claim 19; Fig 1; 60pp; English.	
XX	XX	The present sequence is rice SPPI-related transcription factor #4.	
CC	CC	The SPPI-related transcription factor is useful for transforming a cell	
CC	CC	by introducing SPPI-related transcription factor into a cell. It is also	
CC	CC	useful for producing a transgenic plant by transforming a plant cell with	
CC	CC	SPPI-related transcription factor and regenerating a plant from the	
CC	CC	transformed plant cell. It is also useful to create transgenic plants in	
CC	CC	which SPPI-related transcription factor DNA is present at higher or lower	
CC	CC	levels than normal or in cell types or developmental stages in which they	
CC	CC	are not normally found. The SPPI-related transcription factor DNA is	
CC	CC	useful to prepare antibodies. It is also used as probes for genetically	
CC	CC	and physically mapping the genes that they are a part of, and used as	
CC	CC	markers for traits linked to these genes. Such information is useful in	
CC	CC	plant breeding in order to develop lines with desired phenotype.	
XX	XX	Sequence 487 AA;	
XX	XX	Query Match 42.5%; Score 1295; DB 22; Length 487;	
XX	XX	Best Local Similarity 49.1%; Pred. No. 5.8e-84;	
XX	XX	Matches 291; Conservative 54; Mismatches 124; Indels 124; Gaps 18	

PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 05-MAY-1999; 99US-0132484.
PR 06-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139457.
PR 17-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140921.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0129845.
PR 19-JUL-1999; 99US-0130077.
PR 20-JUL-1999; 99US-0130449.
PR 20-JUL-1999; 99US-0130510.
PR 21-JUL-1999; 99US-0130891.
PR 21-JUL-1999; 99US-0131449.
PR 21-JUL-1999; 99US-0132048.
PR 21-JUL-1999; 99US-0132407.
PR 22-JUL-1999; 99US-0132484.
PR 22-JUL-1999; 99US-0132485.
PR 22-JUL-1999; 99US-0132486.
PR 23-JUL-1999; 99US-0132487.
PR 23-JUL-1999; 99US-0132863.
PR 23-JUL-1999; 99US-0134256.
PR 28-JUL-1999; 99US-0134218.
PR 27-JUL-1999; 99US-0134219.
PR 27-JUL-1999; 99US-0134221.
PR 27-JUL-1999; 99US-0134370.
PR 28-JUL-1999; 99US-0134768.
PR 02-AUG-1999; 99US-0134941.
PR 02-AUG-1999; 99US-0135124.
PR 02-AUG-1999; 99US-0135353.
PR 02-AUG-1999; 99US-0135629.
PR 03-AUG-1999; 99US-0136021.
PR 04-AUG-1999; 99US-0136392.
PR 05-AUG-1999; 99US-0136782.
PR 06-AUG-1999; 99US-0137222.
PR 06-AUG-1999; 99US-0137528.
PR 09-AUG-1999; 99US-0137502.
PR 09-AUG-1999; 99US-0137724.
PR 10-AUG-1999; 99US-0138094.
PR 10-AUG-1999; 99US-0138540.
PR 11-AUG-1999; 99US-0138847.
PR 12-AUG-1999; 99US-0139119.
PR 13-AUG-1999; 99US-0139452.
PR 13-AUG-1999; 99US-0139453.
PR 16-AUG-1999; 99US-0139457.
PR 16-AUG-1999; 99US-0139458.
PR 17-AUG-1999; 99US-0139459.
PR 17-AUG-1999; 99US-0139460.
PR 17-AUG-1999; 99US-0139461.
PR 18-AUG-1999; 99US-0139462.
PR 18-AUG-1999; 99US-0139463.
PR 18-AUG-1999; 99US-0139750.
PR 18-AUG-1999; 99US-0139763.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 28-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.

PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 37.4%; Score 1139.5; DB 21; Length 512;
Best Local Similarity 44.7%; Pred. No. 7.8e-73;
Matches 254; Conservative 80; Mismatches 144; Indels 90; Gaps 15;

QY 40 LONKPPQ-----GGLSERTGS-----GVPKFKSTPPPLPLSPPIPSYFAIP 85
Db 1 MDNSTRQNMNGSANNWSQQSGRTSTSSLELIPKFRSPAPSSISPSLSVSPSTCF--- 57

QY 86 PGLSPAEILDSPVLLNS--NILSPPTGAPVAOSFNWKSNGNQIVKEEDKSPN--- 141
Db 58 ---SPSLFLDSPAFVSSANVLASPTTGALI-----TNVTNRGINEGDKNNNNFN 106

QY 142 ---FSQTRSGPPASSTATVQSQNVTVQTPQWFSFOEATKQDNFSSGKGMKMTENSSMQ 198
Db 107 LPDFSPHTQSGVSAPTTTTTTTTTTTNS--SIFQSQEQKKNQSQEQWSTETPRNQ 164

QY 199 SFSPETASVQTNHNSQSDYGNYPQSQTLRRSDGYNWRKYQKQVKGSENPRSYK 258
Db 165 AVS-----YNGRE-----QRKGEDGYNWRKYQKQVKGSENPRSYK 201

QY 259 CTYPNCPKTKKVRSLDQTEIVYKGTNHPQNTNRNNSNSSI-----AIPHSN 311
Db 202 CTFPNCPKTKKVRSLDQTEIVYKGTNHPQNTNRNNSNSSI-----AIPHSN 261

QY 312 SIRTEIPDQSYATHGSGQMSAATPENSSTISIGDDDFEQSSQCKSGDEYDEPDAXR 371
Db 262 QASDQPNNSNPHQSDQFQMQEDNTTSDVGDDEFEQGS--SIVRDEEDCGSEPEAR 320

QY 372 WKTEGENEGMSPGSRVREPRVVVQTSIDILDDGYRWRKYQKQVKGNNPNPRSYK 431
Db 321 WKGDNETNGGSGSKTVREPRVVVQTSIDILDDGYRWRKYQKQVKGNNPNPRSYK 380

QY 432 THPGCPVKHVERASHDLRAVITTYEGKHNHVPAAAGSGSHSVNRPMPNNAHNHTAA 491
Db 381 TTIGCPVKHVERASHDMRAVITTYEGKHNHVPAAAGSG--YATNRAPOQDSSVPTRPA 439

QY 492 TSVRLLPVHQSDNSLQNSQAPPEQSPPTLEMLQ-----SPGSCFGSGFQPMQSYVN 547
Db 440 IA-----GHSNVTTSQAPYTLQMLHNNNTGTGPGYAMNNNNNSNLQ 483

QY 548 QQQLSDNVFSRTRKEPRDD--MFLSLL 574
Db 484 TQNFVCGGFSRAKEEPEETSFPDPM 511

ID XX AAG31844 standard; Protein; 503 AA.
AC XX AAG31844;
DT XX 17-OCT-2000 (first entry)
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 38310.
XX KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX OS Arabidopsis thaliana.
XX PN EP1033405-A2.
XX PD 06-SEP-2000.
XX PF 25-FEB-2000; 2000EP-0301439.
XX PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.

Db 165 RKGEDGVNWRKYGQKQVKGSENPRSYKCTFPNCPTKKKVERSLBQGITVYVKGSHNP 224
 QY 291 KPQNRTRNSNSSSL-----AIPHSNIRTEIPDOSYATHGSGQMDSAATPENSIS 343
 Db 225 KPQSTRSSSSSTFHSVAVNASLDHNRQASSQPNNSFHSQDSFGMQOEDNTTSDV 284
 QY 344 GDDDFQSGQKCKSGGDEYDEPDADKRWKIEGENEGMSAPGSRTRVREPRVVQTTSDID 403
 Db 285 GDDEFEGGS-SIVSRDEDCGSEPAKRWKGDNETNGGNGGSKTVREPRIVVQTTSDID 343
 QY 404 ILDDGVWRKYGQKQVKGSENPRSYKCTHPGCPVRKHVERASHDLRAVITTYEGKHND 463
 Db 344 ILDDGVWRKYGQKQVKGSENPRSYKCTTIGCPVRKHVERASHDLRAVITTYEGKHND 403
 QY 464 VPAARGSGSHSVNRPMPNNAHTNTAATSVRLLPVHQSDNSLQNSQAPPEGOSPT 523
 Db 404 VPAARGSG-VATNRAPODSSVPIRPAIA-----GHSNYTSSQAPYT 446
 QY 524 LEMLO-----SPGFGFGFGNPMQSYVNOQQLSDNVFSSRTKEPRDD-MFLESLL 574
 Db 447 LQMLHNNNTGTPFGYAMNNNNNSLQTNQNFVGGGFSRAKEPNEBTSFFDSFM 502

RESULT 6

AAE05090
 ID AAE05090 standard; Protein; 488 AA.

AC AAE05090;

DT 12-SEP-2001 (first entry)

DE Rice SPFI-related transcription factor #3.

KW Rice; SPFI-related transcription factor; transgenic plant;
 KW genetical mapping; physical mapping; plant breeding.

OS Oryza sativa.

PN WO200149840-A2.

PD 12-JUL-2001.

PF 22-DEC-2000; 2000WO-US35310.

PR 04-JAN-2000; 2000US-0174325.

PA (DUPO) DU PONT DE NEMOURS & CO E I.

PI Zhu Q, Famodu OO, Rafalski JA;

PP WPI; 2001-441876/47.

DR N-PSDB; AAD09827.

XX Novel isolated SPF-1 related transcription factor polypeptides and
 PT polynucleotides useful for producing transgenic plants -

XX Claim 19; Fig 1; 6opp; English.

CC The present sequence is rice SPFI-related transcription factor #3.
 CC The SPFI-related transcription factor is useful for transforming a cell
 CC by introducing SPFI-related transcription factor into a cell. It is also
 CC useful for producing a transgenic plant by transforming a plant cell with
 CC SPFI-related transcription factor and regenerating a plant from the
 CC transformed plant cell. It is also useful to create transgenic plants in
 CC which SPFI-related transcription factor DNA is present at higher or lower
 CC levels than normal or in cell types or developmental stages in which they
 CC are not normally found. The SPFI-related transcription factor DNA is
 CC useful to prepare antibodies. It is also used as probes for genetically
 CC and physically mapping the genes that they are a part of, and used as
 CC markers for traits linked to these genes. Such information is useful in
 CC plant breeding in order to develop lines with desired phenotype.

XX Sequence 488 AA;

Query Match 36.6%; Score 1117.5; DB 22; Length 488;
 Best Local Similarity 44.6%; Pred. No. 2.7e-71;
 Matches 271; Conservative 54; Mismatches 130; Indels 153; Gaps 22;
 QY 1 MASSSGSLTSASANSFTNFTTSTHPTMTSFS-----DLASPLDNKP---PQGG 49
 Db 1 MASSTGLD-----HGFETTPPFI-TSFTLLSGGGDILLGAGGEERSPRGSRGG 51
 QY 50 LSERTGSGVPKFKSTPPPSLPLSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 109
 Db 52 --ARVGGVPKFKSAQPPSLPLSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 109
 QY 110 TTGAFAVAQFNKWSGSGNQIIVKEEDKSFNFSPQTRSGPPASSTATYQSSNVTVQTOQ 169
 Db 110 TTGAIPAQRYDMKASA---DLIASQDDSRGDFSPHTNSDMAAQAQSPF----- 156
 QY 170 PWSFQATQDNFSSGKGMKMTENSSMSQSFSEIASVQTNHNSGQSDYGNVPPQSOTL 229
 Db 157 --SFKEQEQVVESSKNG-----AAAASSNKSGG---GGN----- 186
 QY 230 SRSDDGVNWRKYGQKQVKGSENPRSYKCTYPCNCTKKKVERSL-DGQITEIVYKGTN 288
 Db 187 -NKLEDGVNWRKYGQKQVKGSENPRSYKCTYNGCMKKKVERSLADGRITQIVYKAHN 245
 QY 289 HPKQNTNRNSNSSSLAIP-----HSNSIRTEIPDQSYATHGSGQMDSAATPEN 338
 Db 246 HPKPLSTA-----ATPLPAPPPAPTTSRRPARAATSTPPRPTPPSPSATRPTTHR 300
 QY 339 SSISIGDDDFEQQSKCKSGGDEYDEPDADKRWKIEGENEGM-SAPGSRTRVREPRVVQ 397
 Db 301 TAA-----RATSPKPSACKEDADN-----EGSGGGMGGAGGNPVRPRLVVQ 343
 QY 398 TTSIDIDLDGVWRKYGQKQVKGSENPRSYKCTHPGCPVRKHVERASHDLRAVITTYE 457
 Db 344 TUSDIDILONGFRWRKYGQKQVKGSENPRSYKCTTVCPCVRKHVERASHDTRAVITTYE 403
 QY 458 GKXNHDPV-----AARGSG-SHSVNRPMNPNASNTHTNTAATSVRLLPVHQSDNSLQ 508
 Db 404 GKXNHDPVPRPRRRTFRGAVAYGWRSGP-----TDVAAAQ----- 442
 QY 509 NORSQAPPEGQSPFTLEMLQSP-GSFGSGFGNPMQSYVNOQQLSDNVFSSRTKEPRDD 567
 Db 443 -----QGPYTLEMLPNPAGLYGCGAG-----AGGAAPPTKDERDD 480
 QY 568 MFLESLLC 575
 Db 481 LFVESLLC 488

RESULT 7

AAE05093

ID AAE05093 standard; Protein; 619 AA.

AC AAE05093;

DT 12-SEP-2001 (first entry)

DE Wheat SPFI-related transcription factor #6.

KW Wheat; SPFI-related transcription factor; transgenic plant;
 KW genetical mapping; physical mapping; plant breeding.

OS Triticum aestivum.

PN WO200149840-A2.

PD 12-JUL-2001.

PF 22-DEC-2000; 2000WO-US35310.

PR 04-JAN-2000; 2000US-0174325.

XX

PA (DUPO) DU PONT DE NEMOURS & CO E I.
XX
PI Zhu Q, Famodu OO, Rafalski JA;
XX
DR WPI; 2001-441876/47.
DR N-PSDB; AAD09830.
XX
XX
PT Novel isolated SPF-1 related transcription factor polypeptides and
PT polynucleotides useful for producing transgenic plants -
XX
XX Claim 19; Page 56-58; 60pp; English.
PS
XX
XX The present sequence is wheat SPFI-related transcription factor #6.
CC The SPFI-related transcription factor is useful for transforming a cell
CC by introducing SPFI-related transcription factor into a cell. It is also
CC useful for producing a transgenic plant by transforming a plant cell with
CC SPFI-related transcription factor and regenerating a plant from the
CC transformed plant cell. It is also useful to create transgenic plants in
CC which SPFI-related transcription factor DNA is present at higher or lower
CC levels than normal or in cell types or developmental stages in which they
CC are not normally found. The SPFI-related transcription factor DNA is
CC useful to prepare antibodies. It is also used as probes for genetically
CC and physically mapping the genes that they are a part of, and used as
CC markers for traits linked to these genes. Such information is useful in
CC plant breeding in order to develop lines with desired phenotype.
XX
XX
SQ Sequence 619 AA;

Query Match 29.7%; Score 904.5; DB 22; Length 619;
Best Local Similarity 39.2%; Pred. No. 6e-56;
Matches 228; Conservative 69; Mismatches 160; Indels 125; Gaps 21;

Qy 41 DNNKPPGGLSERTGVPK-----FKSTPPPSLPLSPSPSSYFA 83
Db 9 DSPNPFSGDLPSAAGSPKPYADRRVAALAGAGARYKAMSPARLPISREPC-----LT 63

Qy 84 IPPCLSPAEILLDSVLLNSNLSPTTGAFVAGSPFNWKS-----SGGNQOIVKEEDK 137
Db 64 IPAGFSALLDSVLLTNKVEPSPTTGLSMAIMHKAHPDILSPDKIRAHEDG 123

Qy 138 SFSNFSF-----QTRSGPPASSTATYQSSNVTVQTQOPWS 172
Db 124 GSRDFEFKPLHNSSSQSLAPMSDLKKHSHMQNSMNPSS-----SSNWNENRPPCS 178

Qy 173 FOEATKODNFSCK--GMM-KTENSMSQSPFPIASVQTNHNSGFSQDYGNYPPQSOT 228
Db 179 -RESSLTNVNSAPNPQVGMVGLTDNMPA-----EVTSEFPQOMN--SSDNAMQEPQSEN 229

Qy 229 LSRRS-DDGYNWRKYGQKVGKSNPRSYKCTYPNCTPKKKKVERSLDGOITETVYKGTN 287
Db 230 VADKSADDGYNWRKYGQKVGKSNPRSYKCTHPNCEVKKLERAVDGLITEVYKGRH 289

Qy 288 NHKPPQNTRRNSNSSSLAIPHNSIRTEIPDOSYAT--HGSGQMDSAATPENS----- 339
Db 290 NHKPPQNTRRLAGG--AVESNQG--EERYDGAAGAAADKSSNALSNLANPNVSPGMVPE 343

Qy 340 -STISIGDDDFEQSQCKSGGDEYDEDEPAKRWKIEGENEGMSAPGSRVTRPRVVQOT 398
Db 344 VPVSVDSDDDIDAGGR-PYPGDDATEEDLESKRRKMSAGIDAALMG-KPNRREPRVVQOT 401

Qy 399 TSDIDILLDDGYRWRKYGQKVGKSNPRSYKCTHPGCPVKRKHVERASHDLRAVITYEG 458
Db 402 VSEVDILLDDGYRWRKYGQKVGKSNPRSYKCTSTGCPVKRKHVERASHDPKSVITYEG 461

Qy 459 KHNHVDPAARGSGSHSVNRPMPNNAHTNTAATSVRLLPVTHQSDNSL----- 507
Db 462 KHNHEVPAAR-NATHENSAPPMKN-----VHQINSMPSSIGGMRRAC 504

Qy 508 -----QNRQOAPPEGQSPFTLEMLQSPGSGFGGFGNPMQS 544
Db 505 EARNFNSQYQSAEETDNLVSLDLGVGISP---NHSDATNQMQS 543

RESULT 8
ABG76918
ID ABG76918 standard; Protein; 278 AA.
XX
AC ABG76918;
XX
DT 05-NOV-2002 (first entry)
XX
DE Wheat WRKY polypeptide #1.
XX
KW WRKY; plant; zinc-finger-type factor; WRKY; SAR; sunflower;
KW SA-dependent structure-activity analysis response; pathogen resistance;
KW maize; wheat; rice; soybean.
XX
OS Triticum aestivum.
XX
PN US2002076775-A1.
XX
PD 20-JUN-2002.
XX
PF 16-MAR-2001; 2001US-0810264.
XX
PR 17-MAR-2000; 2000US-190467P.
XX
PA (CRAN/) CRANE V C.
PA (FAMO/) FAMODU O.
PA (HUX/) HU X.
PA (LUGG/) LU G.
PA (ZHAN/) ZHANG L.
XX
PI Crane VC, Famodu O, Hu X, Lu G, Zhang L;
XX
DR WPI; 2002-582922/62.
DR N-PSDB; ABS59414.
XX
XX New isolated plant-specific zinc-finger-type factor polynucleotide,
PT useful for e.g. regulating the SA-dependent structure-activity analysis
PT response in a plant -
XX
XX Claim 9; Page 51-52; 66pp; English.
XX
XX The invention relates to an isolated polynucleotide encoding a
CC plant-specific zinc-finger-type factor (WRKY) protein. The DNA and
CC protein are used to modulate the level of a WRKY protein in a plant and
CC to regulate the SA-dependent structure-activity analysis (SAR) response
CC in a plant. The sequences can be used to engineer plants to resist
CC pathogens such as viruses, bacteria, insects and fungi, and to survive
CC stresses. They may also be used as a molecular probe to track inheritance
CC of corresponding loci in genetic crosses and facilitate the plant
CC breeding process, to isolate, identify and genetically map WRKY and other
CC closely related disease resistance genes and to find genes and their
CC promoters that respond to a WRKY domain. This sequence represents a WRKY
CC polypeptide of the invention.
XX
XX Sequence 278 AA;

Query Match 29.1%; Score 888.5; DB 23; Length 278;
Best Local Similarity 52.4%; Pred. No. 2.7e-55;
Matches 184; Conservative 41; Mismatches 47; Indels 79; Gaps 11;

Qy 230 SRRSDGYNWRKYGQKVGKSNPRSYKCTYPNCTPKKKKVERSLDGOITETVYKGTN 288
Db 2 NKQVEDGYNWRKYGQKVGKSNPRSYKCTYNNCSMKKKVERSLADGRITQIVYKGAHD 61

Qy 289 HPKPQNTRRNSNSSS-LAIPHNSIRTEIPDOSYATHGSGQMDSAATPENSISITGDD 347
Db 62 HPKPPTRRNSSCAAVIAEDHTN-----GSEHSGPTPENSIVTFGDDE 105

Qy 348 FEQSSQCKSGGDEYDEDEPAKRWKIEGENEGMS-APGS--RTVREPRVVVQTTSDIDI 404
Db 106 ADNGA-----EPETKRKRKEHGDNEGSGGTGACVKVPRPLVUOTLSDIDI 152

Qy 405 LDDGYRWRKYGQKVGKSNPRSYKCTHPGCPVKRKHVERASHDLRAVITYEGKHNHDV 464


```
Db      153 LDDGFRWRKYGQVKGNPNRYSYKCTTVGCPVRKHVERASHDNRAVITTYEGRSHDV 212
Qy      465 PAARGSGSHSVNRPMPNNAASHNTAATSVRLPLVHQSDNSLQNRSQAPPGSQSPFTL 524
Db      213 PVGRGAGA---SRALPTSSSD-----SSVVVCPA-----AAGQAPYTL 248
Qy      525 EMLQSPGSGFSGFGNPMQSYVNVQQLSDNVFSSRTKEPRDDMFLESLIC 575
Db      249 EMLANPAA-GRGYA-----AKDEPRDDMFVESLIC 278

RESULT 9
AAG46847
ID      AAG46847 standard; Protein; 309 AA.
XX
AC      AAG46847;
XX
DT      18-OCT-2000 (first entry)
DE      Arabidopsis thaliana protein fragment SEQ ID NO: 58980.
XX
KW      Protein identification; signal transduction pathway; metabolic pathway;
KW      hybridisation assay; genetic mapping; gene expression control; promoter;
KW      termination sequence.
XX
OS      Arabidopsis thaliana.
XX
PN      EF1033405-A2.
XX
PD      06-SEP-2000.
XX
PF      25-FEB-2000; 2000EP-0301439.
XX
PR      25-FEB-1999; 99US-0121825.
PR      05-MAR-1999; 99US-0123180.
PR      09-MAR-1999; 99US-0123548.
PR      23-MAR-1999; 99US-0125788.
PR      25-MAR-1999; 99US-0126264.
PR      29-MAR-1999; 99US-0126785.
PR      01-APR-1999; 99US-0127462.
PR      06-APR-1999; 99US-0128234.
PR      08-APR-1999; 99US-0128714.
PR      16-APR-1999; 99US-0129845.
PR      19-APR-1999; 99US-0130077.
PR      21-APR-1999; 99US-0130449.
PR      23-APR-1999; 99US-0130510.
PR      28-APR-1999; 99US-0130891.
PR      30-APR-1999; 99US-0131449.
PR      30-APR-1999; 99US-0132048.
PR      04-MAY-1999; 99US-0132407.
PR      04-MAY-1999; 99US-0132484.
PR      05-MAY-1999; 99US-0132485.
PR      06-MAY-1999; 99US-0132486.
PR      06-MAY-1999; 99US-0132487.
PR      07-MAY-1999; 99US-0132863.
PR      11-MAY-1999; 99US-0134256.
PR      14-MAY-1999; 99US-0134218.
PR      14-MAY-1999; 99US-0134219.
PR      14-MAY-1999; 99US-0134221.
PR      14-MAY-1999; 99US-0134370.
PR      18-MAY-1999; 99US-0134768.
PR      19-MAY-1999; 99US-0134941.
PR      20-MAY-1999; 99US-0135184.
PR      21-MAY-1999; 99US-0135353.
PR      24-MAY-1999; 99US-0135629.
PR      25-MAY-1999; 99US-0136021.
PR      27-MAY-1999; 99US-0136392.
PR      28-MAY-1999; 99US-0136782.
PR      01-JUN-1999; 99US-0137222.
PR      03-JUN-1999; 99US-0137528.
PR      04-JUN-1999; 99US-0137502.
PR      07-JUN-1999; 99US-0137724.
PR      08-JUN-1999;
PR      10-JUN-1999; 99US-0138094.
PR      10-JUN-1999; 99US-0138540.
PR      14-JUN-1999; 99US-0138847.
PR      16-JUN-1999; 99US-0139119.
PR      16-JUN-1999; 99US-0139452.
PR      16-JUN-1999; 99US-0139453.
PR      17-JUN-1999; 99US-0139492.
PR      18-JUN-1999; 99US-0139454.
PR      18-JUN-1999; 99US-0139455.
PR      18-JUN-1999; 99US-0139456.
PR      18-JUN-1999; 99US-0139457.
PR      18-JUN-1999; 99US-0139458.
PR      18-JUN-1999; 99US-0139459.
PR      18-JUN-1999; 99US-0139460.
PR      18-JUN-1999; 99US-0139461.
PR      18-JUN-1999; 99US-0139462.
PR      18-JUN-1999; 99US-0139463.
PR      18-JUN-1999; 99US-0139750.
PR      18-JUN-1999; 99US-0139763.
PR      21-JUN-1999; 99US-0139817.
PR      22-JUN-1999; 99US-0139899.
PR      23-JUN-1999; 99US-0140353.
PR      23-JUN-1999; 99US-0140354.
PR      24-JUN-1999; 99US-0140695.
PR      28-JUN-1999; 99US-0140823.
PR      29-JUN-1999; 99US-0140991.
PR      30-JUN-1999; 99US-0141287.
PR      01-JUL-1999; 99US-0141842.
PR      01-JUL-1999; 99US-0142154.
PR      02-JUL-1999; 99US-0142055.
PR      06-JUL-1999; 99US-0142390.
PR      08-JUL-1999; 99US-0142803.
PR      09-JUL-1999; 99US-0142920.
PR      12-JUL-1999; 99US-0142977.
PR      13-JUL-1999; 99US-0143542.
PR      14-JUL-1999; 99US-0143624.
PR      15-JUL-1999; 99US-0144005.
PR      16-JUL-1999; 99US-0144085.
PR      16-JUL-1999; 99US-0144086.
PR      19-JUL-1999; 99US-0144325.
PR      19-JUL-1999; 99US-0144331.
PR      19-JUL-1999; 99US-0144332.
PR      19-JUL-1999; 99US-0144333.
PR      19-JUL-1999; 99US-0144334.
PR      19-JUL-1999; 99US-0144335.
PR      20-JUL-1999; 99US-0144352.
PR      20-JUL-1999; 99US-0144632.
PR      20-JUL-1999; 99US-0144884.
PR      21-JUL-1999; 99US-0144814.
PR      21-JUL-1999; 99US-0145086.
PR      21-JUL-1999; 99US-0145088.
PR      22-JUL-1999; 99US-0145085.
PR      22-JUL-1999; 99US-0145087.
PR      22-JUL-1999; 99US-0145089.
PR      22-JUL-1999; 99US-0145192.
PR      23-JUL-1999; 99US-0145145.
PR      23-JUL-1999; 99US-0145218.
PR      23-JUL-1999; 99US-0145224.
PR      26-JUL-1999; 99US-0145276.
PR      27-JUL-1999; 99US-0145913.
PR      27-JUL-1999; 99US-0145918.
PR      27-JUL-1999; 99US-0145919.
PR      28-JUL-1999; 99US-0145951.
PR      02-AUG-1999; 99US-0146386.
PR      02-AUG-1999; 99US-0146388.
PR      02-AUG-1999; 99US-0146389.
PR      03-AUG-1999; 99US-0147038.
PR      04-AUG-1999; 99US-0147204.
PR      04-AUG-1999; 99US-0147302.
PR      05-AUG-1999; 99US-0147192.
PR      05-AUG-1999; 99US-0147260.
PR      06-AUG-1999; 99US-0147303.
PR      06-AUG-1999; 99US-0147416.
```

PR	09-AUG-1999;	99US-0147493.	QY	58	VPKFKSTPPPSLPLGPPPISSSYFAIPGLSPAELLDSPVLNNSNILPSPPTGAFVAQ	117
PR	09-AUG-1999;	99US-0147935.	Db	10	VPKFKTATP-----SPLPLSPSPYFTMPGLTADFLDPLFTSSNILPSPPTGTFPAQ	64
PR	10-AUG-1999;	99US-0148171.	QY	118	SNWKSSSCG---NQOIVKEEDKSNFSPFQTRSGPPASSTATYQSSNVTVTQOPWSFQ	174
PR	11-AUG-1999;	99US-0148319.	Db	65	SLNY--NNGLLIDKNEIKYEDTT-----PP-----LFUPSMVT---QPLPOL	102
PR	12-AUG-1999;	99US-0148565.	QY	175	EATKQDNFSSGKMMKNTSSSMQSPFSAIVQTNHNGFQSDYGNYPQSQTLSSRRSD	234
PR	13-AUG-1999;	99US-0148684.	Db	103	DLFKSEIMSSNK-----	117
PR	16-AUG-1999;	99US-0149358.	QY	235	DGYNMRKYQKQVKGSENPRSYKCTYPCNCTKKKVERSL-DQQTETIYVKGHNHPKQ	293
PR	17-AUG-1999;	99US-0149175.	Db	118	DGYNMRKYQKQVKGSENPRSYKCTYPCNCTKKKVERSL-DQQTETIYVKGHNHPKQ	177
PR	18-AUG-1999;	99US-0149426.	QY	294	NTRNRSNSSLAIHPSNIRTEIPDQSYATHGSGQMDSAATPENSISIGDDDFQSSQ	353
PR	20-AUG-1999;	99US-0149723.	Db	178	STKRSSTAIA---AHQNS-----SNGD-----	197
PR	20-AUG-1999;	99US-0149929.	QY	354	KCKSGGDEYDEDEPDAKRWKIEGENEGMSAPGSRTVREPRVVVQTTSDIDLDGGRWRK	413
PR	23-AUG-1999;	99US-0149902.	Db	198	-----GKDIGEDETEAKRWRE-EN-----VKERVVVQTTSDIDLDGGRWRK	241
PR	23-AUG-1999;	99US-0149930.	QY	414	YGQVVKGNPNRPSYKCTHPCGCPVKRKHVERASHDLRAVITTYEGKHNDVPAAR	468
PR	25-AUG-1999;	99US-0150566.	Db	242	YGQVVKGNPNRPSYKCTHPCGCPVKRKHVERASHDLRAVITTYEGKHNDVPAAR	296
PR	26-AUG-1999;	99US-0150884.	QY	178	STKRSSTAIA---AHQNS-----SNGD-----	197
PR	27-AUG-1999;	99US-0151066.	Db	354	KCKSGGDEYDEDEPDAKRWKIEGENEGMSAPGSRTVREPRVVVQTTSDIDLDGGRWRK	413
PR	27-AUG-1999;	99US-0151080.	QY	198	-----GKDIGEDETEAKRWRE-EN-----VKERVVVQTTSDIDLDGGRWRK	241
PR	30-AUG-1999;	99US-0151303.	Db	414	YGQVVKGNPNRPSYKCTHPCGCPVKRKHVERASHDLRAVITTYEGKHNDVPAAR	468
PR	31-AUG-1999;	99US-0151438.	QY	242	YGQVVKGNPNRPSYKCTHPCGCPVKRKHVERASHDLRAVITTYEGKHNDVPAAR	296
PR	01-SEP-1999;	99US-0151930.	Db	178	STKRSSTAIA---AHQNS-----SNGD-----	197
PR	07-SEP-1999;	99US-0152363.	QY	354	KCKSGGDEYDEDEPDAKRWKIEGENEGMSAPGSRTVREPRVVVQTTSDIDLDGGRWRK	413
PR	10-SEP-1999;	99US-0153070.	Db	198	-----GKDIGEDETEAKRWRE-EN-----VKERVVVQTTSDIDLDGGRWRK	241
PR	13-SEP-1999;	99US-0153758.	QY	414	YGQVVKGNPNRPSYKCTHPCGCPVKRKHVERASHDLRAVITTYEGKHNDVPAAR	468
PR	15-SEP-1999;	99US-0154018.	Db	242	YGQVVKGNPNRPSYKCTHPCGCPVKRKHVERASHDLRAVITTYEGKHNDVPAAR	296
PR	16-SEP-1999;	99US-0154039.	QY	178	STKRSSTAIA---AHQNS-----SNGD-----	197
PR	20-SEP-1999;	99US-0154779.	Db	354	KCKSGGDEYDEDEPDAKRWKIEGENEGMSAPGSRTVREPRVVVQTTSDIDLDGGRWRK	413
PR	22-SEP-1999;	99US-0155139.	QY	198	-----GKDIGEDETEAKRWRE-EN-----VKERVVVQTTSDIDLDGGRWRK	241
PR	23-SEP-1999;	99US-0155486.	Db	414	YGQVVKGNPNRPSYKCTHPCGCPVKRKHVERASHDLRAVITTYEGKHNDVPAAR	468
PR	24-SEP-1999;	99US-0155659.	QY	242	YGQVVKGNPNRPSYKCTHPCGCPVKRKHVERASHDLRAVITTYEGKHNDVPAAR	296
PR	28-SEP-1999;	99US-0156458.	Db	178	STKRSSTAIA---AHQNS-----SNGD-----	197
PR	29-SEP-1999;	99US-0156596.	QY	354	KCKSGGDEYDEDEPDAKRWKIEGENEGMSAPGSRTVREPRVVVQTTSDIDLDGGRWRK	413
PR	04-OCT-1999;	99US-0157117.	Db	198	-----GKDIGEDETEAKRWRE-EN-----VKERVVVQTTSDIDLDGGRWRK	241
PR	06-OCT-1999;	99US-0157753.	QY	414	YGQVVKGNPNRPSYKCTHPCGCPVKRKHVERASHDLRAVITTYEGKHNDVPAAR	468
PR	07-OCT-1999;	99US-0158029.	Db	242	YGQVVKGNPNRPSYKCTHPCGCPVKRKHVERASHDLRAVITTYEGKHNDVPAAR	296
PR	08-OCT-1999;	99US-0158232.	QY	178	STKRSSTAIA---AHQNS-----SNGD-----	197
PR	12-OCT-1999;	99US-0158369.	Db	354	KCKSGGDEYDEDEPDAKRWKIEGENEGMSAPGSRTVREPRVVVQTTSDIDLDGGRWRK	413
PR	13-OCT-1999;	99US-0159293.	QY	198	-----GKDIGEDETEAKRWRE-EN-----VKERVVVQTTSDIDLDGGRWRK	241
PR	13-OCT-1999;	99US-0159294.	Db	414	YGQVVKGNPNRPSYKCTHPCGCPVKRKHVERASHDLRAVITTYEGKHNDVPAAR	468
PR	13-OCT-1999;	99US-0159295.	QY	242	YGQVVKGNPNRPSYKCTHPCGCPVKRKHVERASHDLRAVITTYEGKHNDVPAAR	296
PR	14-OCT-1999;	99US-0159329.	Db	178	STKRSSTAIA---AHQNS-----SNGD-----	197
PR	14-OCT-1999;	99US-0159330.	QY	354	KCKSGGDEYDEDEPDAKRWKIEGENEGMSAPGSRTVREPRVVVQTTSDIDLDGGRWRK	413
PR	14-OCT-1999;	99US-0159331.	Db	198	-----GKDIGEDETEAKRWRE-EN-----VKERVVVQTTSDIDLDGGRWRK	241
PR	14-OCT-1999;	99US-0159637.	QY	414	YGQVVKGNPNRPSYKCTHPCGCPVKRKHVERASHDLRAVITTYEGKHNDVPAAR	468
PR	14-OCT-1999;	99US-0159638.	Db	242	YGQVVKGNPNRPSYKCTHPCGCPVKRKHVERASHDLRAVITTYEGKHNDVPAAR	296
PR	18-OCT-1999;	99US-0159584.	QY	178	STKRSSTAIA---AHQNS-----SNGD-----	197
PR	21-OCT-1999;	99US-0160741.	Db	354	KCKSGGDEYDEDEPDAKRWKIEGENEGMSAPGSRTVREPRVVVQTTSDIDLDGGRWRK	413
PR	21-OCT-1999;	99US-0160767.	QY	198	-----GKDIGEDETEAKRWRE-EN-----VKERVVVQTTSDIDLDGGRWRK	241
PR	21-OCT-1999;	99US-0160768.	Db	414	YGQVVKGNPNRPSYKCTHPCGCPVKRKHVERASHDLRAVITTYEGKHNDVPAAR	468
PR	21-OCT-1999;	99US-0160770.	QY	242	YGQVVKGNPNRPSYKCTHPCGCPVKRKHVERASHDLRAVITTYEGKHNDVPAAR	296
PR	21-OCT-1999;	99US-0160814.	Db	178	STKRSSTAIA---AHQNS-----SNGD-----	197
PR	21-OCT-1999;	99US-0160815.	QY	354	KCKSGGDEYDEDEPDAKRWKIEGENEGMSAPGSRTVREPRVVVQTTSDIDLDGGRWRK	413
PR	22-OCT-1999;	99US-0160980.	Db	198	-----GKDIGEDETEAKRWRE-EN-----VKERVVVQTTSDIDLDGGRWRK	241
PR	22-OCT-1999;	99US-0160981.	QY	414	YGQVVKGNPNRPSYKCTHPCGCPVKRKHVERASHDLRAVITTYEGKHNDVPAAR	468
PR	22-OCT-1999;	99US-0160989.	Db	242	YGQVVKGNPNRPSYKCTHPCGCPVKRKHVERASHDLRAVITTYEGKHNDVPAAR	296
PR	25-OCT-1999;	99US-0161404.	QY	178	STKRSSTAIA---AHQNS-----SNGD-----	197
PR	25-OCT-1999;	99US-0161405.	Db	354	KCKSGGDEYDEDEPDAKRWKIEGENEGMSAPGSRTVREPRVVVQTTSDIDLDGGRWRK	413
PR	25-OCT-1999;	99US-0161406.	QY	198	-----GKDIGEDETEAKRWRE-EN-----VKERVVVQTTSDIDLDGGRWRK	241
PR	26-OCT-1999;	99US-0161359.	Db	414	YGQVVKGNPNRPSYKCTHPCGCPVKRKHVERASHDLRAVITTYEGKHNDVPAAR	468
PR	26-OCT-1999;	99US-0161360.	QY	242	YGQVVKGNPNRPSYKCTHPCGCPVKRKHVERASHDLRAVITTYEGKHNDVPAAR	296
PR	26-OCT-1999;	99US-0161361.	Db	178	STKRSSTAIA---AHQNS-----SNGD-----	197
PR	28-OCT-1999;	99US-0161920.	QY	354	KCKSGGDEYDEDEPDAKRWKIEGENEGMSAPGSRTVREPRVVVQTTSDIDLDGGRWRK	413
PR	28-OCT-1999;	99US-0161992.	Db	198	-----GKDIGEDETEAKRWRE-EN-----VKERVVVQTTSDIDLDGGRWRK	241
PR	28-OCT-1999;	99US-0161993.	QY	414	YGQVVKGNPNRPSYKCTHPCGCPVKRKHVERASHDLRAVITTYEGKHNDVPAAR	468
PR	29-OCT-1999;	99US-0162142.	Db	242	YGQVVKGNPNRPSYKCTHPCGCPVKRKHVERASHDLRAVITTYEGKHNDVPAAR	296

Query Match 29.0%; Score 884; DB 21; Length 309;
Best Local Similarity 47.7%; Pred. No. 6.5e-55;
Matches 198; Conservative 30; Mismatches 55; Indels 132; Gaps 13;

RESULT 10

ABP81179
ID ABP81179 standard; Protein; 514 AA.
XX ABP81179;
AC ABP81179;
XX
DT 27-FEB-2003 (first entry)
XX
XX Arabidopsis thaliana protein #7 modulated by PTGS.
DE Posttranscriptional gene silencing; PTGS; plant; transformation.
XX
XX Arabidopsis thaliana.
OS
XX
XX W0200281695-A2.
XX
XX 17-OCT-2002.
PD
XX
XX 05-APR-2002; 2002WO-EP03806.
PF
XX
XX 06-APR-2001; 2001US-282049P.
PR
XX
XX (SYGN) SYNGENTA PARTICIPATIONS AG.
PA
XX (FRIE-) FRIEDRICH MIESCHER INST.
XX
XX Zhu T, Glazov EA, Meins F, Wang X, Chang H;
XX
XX WPI; 2003-103337/09.
DR
XX N-PSDB; AB242023.

Novel polynucleic acid segment useful for modulating gene expression within a cell by posttranscriptional gene silencing, and for augmenting a plant cell genome -
PS Claim 53; Page 164-166; 438pp; English.
XX
XX The invention relates to a novel isolated polynucleic acid segment modulated within a cell by posttranscriptional gene silencing (PTGS). The invention specifically relates to a method to identify an expression product that is modulated by PTGS. The polynucleotide is useful for modulating the gene expression within a cell by PTGS, by introducing the polynucleic acid into a cell and expressing the nucleic acid segment in the cell to form a product. The polynucleic acid segment is also useful for augmenting a cell genome, and for augmenting a plant genome, by

Db 464 HDMAGPA--SASGQTRVRPESDTISLDGMGISPAEAENTNSQGRMLSEFGDSQIHTS 521
 QY 504 DNSLQNRQAPPEGQSPFTLEMLQSPGFGSGFNGPMQSYVYNQQLSDNVFSSRTKEE 563
 Db 522 NSNFK-----FVHTTAPGYFV-----LNNNSPYGSKEN 552
 QY 564 PRD 566
 Db 553 PSD 555

RESULT 12
 AAU92971
 ID AAU92971 standard; Protein; 687 AA.
 XX AAU92971;
 AC AAU92971;
 DT 02-JUL-2002 (first entry)
 XX Arabidopsis transcription factor #9.
 DE Arabidopsis transcription factor #9.
 XX Agriculture; metabolic chemical; environmental stress; drought;
 KW microbial disease resistance; herbicide resistance; seed yield;
 KW fruit yield; growth rate; leaf senescence; flower senescence.
 KW plant; transcription factor; transgenic.
 XX Arabidopsis thaliana.
 OS Arabidopsis thaliana.
 XX WO200215675-A1.
 PN 28-FEB-2002.
 PD 22-AUG-2001; 2001WO-US26189.
 PF 22-AUG-2001; 2000US-227439P.
 PR 16-NOV-2000; 2000US-071399A.
 PR 16-APR-2001; 2001US-083794A.
 XX (MEND-) MENDEL BIOTECHNOLOGY INC.
 PA (PILG/) PILGRIM M.
 PA (CREE/) CREELMAN R.
 PA (DUBE/) DUBELL A. J.
 PA (HEAR/) HEARD J.
 PA (JIAN/) JIANG C.
 PA (KEDD/) KEDDIE J.
 PA (ADAM/) ADAM L.
 PA (RATC/) RATCLIFF O.
 PA (REUB/) REUBER J. L.
 PA (RIEC/) RIECHMANN J. L.
 PA (YUGG/) YU G.
 PA (PINE/) PINEDA O.
 XX Pilgrim M, Creelman R, Dubell AJ, Heard J, Jiang C, Keddle J;
 PI Adam L, Ratcliff O, Reuber JL, Riechmann JL, Yu G, Pineda O;
 XX WPI; 2002-292022/33.
 DR N-PSDB; ABK65157.
 DR An isolated or recombinant polynucleotide used to produce a transgenic
 PT plant -
 PT Claim 40; Page 105-108; 941pp; English.
 PS The invention relates to 1 of 232 isolated or recombinant polynucleotides
 CC encoding an Arabidopsis thaliana transcription factor, their variants,
 CC complements, fragments, or related polynucleotide with 31% to 95%
 CC sequence identity, where the plant possesses an altered trait as compared
 CC to a wild-type or reference plant, or the plant exhibits an altered
 CC phenotype as compared to a wild-type or reference plant, or the plant
 CC exhibits ectopic expression or altered expression of one or more genes
 CC associated with a plant trait as compared to a wild plant. Also included
 CC are a transgenic plant comprising the polynucleotides, a computer
 CC readable medium having stored sequence information, and identifying a

CC homologue sequence from a database comprising a plurality of known plant
 CC sequences comprising inputting sequence information selected from one of
 CC 464 fully defined sequences given in the specification. The isolated or
 CC recombinant polynucleotide is used for producing a plant having a
 CC modified trait, the method comprising selecting a polynucleotide that
 CC encodes a polypeptide or an antisense nucleic acid, inserting the
 CC polynucleotide or antisense nucleic acid into an expression vector,
 CC introducing the vector into a plant or a cell of a plant to overexpress
 CC the polypeptide or antisense nucleic acid, thereby producing a modified
 CC plant, and selecting for a modified trait (e.g. increased
 CC production of agriculturally useful proteins or metabolic chemicals,
 CC pest tolerance, environmental stress response (e.g. drought), microbial
 CC disease resistance, herbicide resistance, seed and fruit yield, growth
 CC rate, leaf and flower senescence and many other traits listed in the
 CC specification). The present sequence is one of the 232 proteins which are
 CC A. thaliana transcription factors.

XX SQ Sequence 687 AA;

Query Match 27.9%; Score 852; DB 23; Length 687;
 Best Local Similarity 37.1%; Pred. No. 3.9e-52;
 Matches 229; Conservative 63; Mismatches 146; Indels 180; Gaps 24;
 QY 47 QGGLSE-----RTGSGVPKFKSTPPPSLPLSPPIPSFYAI-----PGLSP 90
 Db 72 RGLSERIAARAGFNAPR-----LNTENIRNTDFSDNSLRSPCLTISPSGLSP 121
 QY 91 AELLSPVLLNSSNLPSPPTTGAFVAQSFNWKSSSGN--QQIVKEE--DKSFSNFSQ 145
 Db 122 ATLLSPVFLSNPLAQPSPTTG-----KFPFLPGVNGNALSEKAKDEFDDIGASFSTH 176
 QY 146 TRSGPPASSTATYQSSNVTVQTPQMSFOEATKQDNFSSGKGMKMTENSMSQSFSEITA 205
 Db 177 ----FVSRSSSF-----FQGTTEMSVDYGNYNRSSHQSAEEVKPGSE 218
 QY 206 SVQTNHSGFQSDYGNYPQSQTL-----SRSD-----DG 237
 Db 219 NTESSNLYGIETD--NQNGQNKTSVTTNTSLTVDHQBEEBEEQRGDSMAGGAPAE 276
 QY 238 NNRKYGQKGVKGSNPRSYKCTYPCNPKKVERSLOGOITEIVYKGTNNHFKPQNTTR 297
 Db 277 NNRKYGQKLVKGSYPRSYKCTNPNQVKKVEREGHTEIITYGAHNLKPPNRR 336
 QY 298 N-----SSNSSSLAIPHNSNIRTEIPDQSYAT 324
 Db 337 SGMQVDGTEQVEQQOQQRDSAAATVWSCNTQQGGSNENN--EGSTRFEYGNQSGSI 393
 QY 325 HG--SQMDS---AATPENSISIGDDDFEASSQCKS-----GGDEYDEDEPAKR 371
 Db 394 QAQTGGQYESGDPVVVVDASSTFSDNEDDRGTHGSVSLGYDGGGGGGGGEGSESKR 453
 QY 372 WKIEGENEGMSAPGSTRVREPRVVVQTTSDIDLDGGRWKYQKVKGNPNPRSYK 431
 Db 454 RKLFAFAEMSG--STRAIREPRVVVQTTSDVDLDGGRWKYQKVKGNPNPRSYK 512
 QY 432 THPGCPVRKHVERASHDLRAVITYYEGKINHDPAAAR-----GSGSHSVNRMPNPNASNHT 487
 Db 513 TAFGCTVRKHVERASHDLKSVITYYEGKINHDPAAARNSSHGCGGDSGN-----GNSG 565
 QY 488 NTAATSVRLLPVTHQSDNSLQNRQSAAPPEG-----QSPFTEMLQSPGSP----- 533
 Db 566 GSAAVS-----HHYHNGHSEPRGRFDRQVTTNNQSPFS-----RPFSPQPHL 609
 QY 534 ----GFSGFGNPMQSYVN 547
 Db 610 GPSPGFS--FELGQTGLVN 626

RESULT 13
 ABP81191
 ID ABP81191 standard; Protein; 513 AA.
 XX ABP81191;
 AC ABP81191;

XX 27-FEB-2003 (first entry)
XX Arabidopsis thaliana protein #19 modulated by PTGS.
XX Posttranscriptional gene silencing; PTGS; plant; transformation.
XX Arabidopsis thaliana.
XX OS Arabidopsis thaliana.
XX WO200281695-A2.
XX 17-OCT-2002.
XX 05-APR-2002; 2002WO-EP03806.
XX 06-APR-2001; 2001US-282049P.
XX (SYGN) SYNGENTA PARTICIPATIONS AG.
XX (FRIE-) FRIEDRICH MIESCHER INST.
XX Zhu T, Glazov EA, Meins F, Wang X, Chang H;
XX WPI; 2003-103337/09.
XX N-ESDB; ABZ42035.
XX Novel polynucleic acid segment useful for modulating gene expression
XX within a cell by posttranscriptional gene silencing, and for augmenting
XX a plant cell genome -
XX Claim 53; Page 192-194; 438pp; English.
XX The invention relates to a novel isolated polynucleic acid segment
XX modulated within a cell by posttranscriptional gene silencing (PTGS). The
XX invention specifically relates to a method to identify an expression
XX product that is modulated by PTGS. The polynucleotide is useful for
XX modulating the gene expression within a cell by PTGS, by introducing the
XX polynucleic acid into a cell and expressing the nucleic acid segment in
XX the cell to form a product. The polynucleic acid segment is also useful
XX for augmenting a cell genome, and for augmenting a plant genome, by
XX contacting a plant cell with the segment to produce a transformed plant
XX cell, and growing the transformed plant cell to produce a differentiated
XX transformed plant. The sequences shown in ABP81173 - ABP81298 represent
XX the product of a segment of A. thaliana cDNA modulated by PTGS.
XX Sequence 513 AA;
SQ
Query Match 27.8%; Score 848.5; DB 24; Length 513;
Best Local Similarity 41.0%; Pred. No. 4.6e-52;
Matches 210; Conservative 52; Mismatches 169; Indels 81; Gaps 15;
QY 5 SGSLDTSASANSFTNFTFTSTPFTTSTFSDLLASPL-----DNNKPPQG 48
Db 36 SGGVGFSGPMTLVSNLFS-DPDEFKFSQLLAGAMASPRAAVAANVATAHQTPVS 94
QY 49 GLSERTGSG---VPKFKSTPPPSLPLSPPTSPSSYFAIPGLSPABELLSPVLLNSSNI 105
Db 95 SVCGGGGSGGVDRFRKQSRPTGLMITQPP---GMFTVPGLSPALLDSPSPFG--- 146
QY 106 LPSPTTGAFVAQSFNWKSSGNGNOQIVKEEDKSFNSFOTRSGPPASSTA-TVQSSNVT 164
Db 147 LFSPLQGTTF-----GMTHQQAQAQVTAQVQGNVH 177
QY 165 VOTQPPWSFQBATKDNFSGKGMKMTENSSSMQSFSPETASVQ-TNHSNGFQSDYG--N 221
Db 178 MQQSCQSEYFSTQOQQOQQOQOQASLTIPFSSAPRQIRASVOETSGQRETSEISVFE 237
QY 222 YPQSQTLRRSDGYNWRKYQKQVKGSENPRTYKCTYPCNCTKKKVERSLEDGQITEI 281
Db 238 HRSQPNADKPADGYNWRKYQKQVKGSDPFRSYKCTHPACPVKKVERSLEDGQITEI 297
QY 282 VYKQTHNHPKPNTRNSSSSSLAIPIHNSIRTEIPDQSYATHGSGQMSAATPENSII 341
Db 298 IYKQCHNHELFPKRGNNNGSCKS-----SDIANQ-FQTSNSSLNKSQRDQETSQV 346

QY 342 SIGDDDFEQSSQK-----CKSGGDEYDEDEPDAKRWKIEGENEGMSAPGSRTVREPRVVVQ 397
Db 347 TTTEQMSSEASDSEVGNABTSVGERHEDEPDPKRRNTEVRVSPVASSSHRTVTPEPRIIVQ 406
QY 398 TTSDIDILDDGYRWRYGQKVKVGNPNRPSYKCTHPGCPVRKHVERASHDLRAVITTYE 457
Db 407 TTSEVDLLDDGYRWRYGQKVKVGNPNRPSYKCTTPDCGVRKHVERAAATDPKAVVTTYE 466
QY 458 GKHNDVPAARGSGSHSVNRPMPNNAHNTNT 489
Db 467 GKHNDVPAARTS-SHQLR---PNNQHN-TST 493
RESULT 14
AAG46848
ID AAG46848 standard; Protein; 279 AA.
XX AAG46848;
XX 18-OCT-2000 (first entry)
XX Arabidopsis thaliana protein fragment SEQ ID NO: 58981.
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX Arabidopsis thaliana.
XX EF1033405-A2.
XX 06-SEP-2000.
XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 28-APR-1999; 99US-0130891.
XX 30-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 04-MAY-1999; 99US-0132407.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 06-MAY-1999; 99US-0132487.
XX 07-MAY-1999; 99US-0132863.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 28-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142330.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145132.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145227.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.

PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 26.8%; Score 818.5; DB 21; Length 279;
Best Local Similarity 47.3%; Pred. No. 2.7e-50;

Db 442 GSSGTLQGLATQTHNHNHVPMPHSRSEGLATA-----NSSLDFDQSH--- 485

Qy 517 EGQSPFTLEMLQSPGSGFGSGFGNPMQSYVNOQQLSD 553

Db 486 -----LRHPTGFSV-----YIGQSELSLSD 503

Search completed: January 20, 2004, 16:19:24
Job time : 49 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 20, 2004, 16:16:52 ; Search time 21 Seconds
(without alignments)
1158.512 Million cell updates/sec

Title: US-09-890-811B-10

Perfect score: 3050

Sequence: 1 MASSGSLDTSASANSFTNF.....FSSRTKEPRDDMFLESLLC 575

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA.*

1: /cgn2_6/ptodata/1/iaa/5A COMB.pap.*

2: /cgn2_6/ptodata/1/iaa/5B COMB.pap.*

3: /cgn2_6/ptodata/1/iaa/6A COMB.pap.*

4: /cgn2_6/ptodata/1/iaa/6B COMB.pap.*

5: /cgn2_6/ptodata/1/iaa/PCTUS COMB.pap.*

6: /cgn2_6/ptodata/1/iaa/backfile1.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	163	5.3	1312	3	US-09-041-886-19
2	163	5.3	1312	4	US-09-848-281-2
3	154	5.0	2842	1	US-07-741-940-7
4	154	5.0	2842	1	US-08-289-548A-7
5	154	5.0	2842	1	US-08-452-654-7
6	154	5.0	2842	4	US-08-449-731-7
7	154	5.0	2843	1	US-07-741-940-2
8	154	5.0	2843	1	US-08-289-548A-2
9	154	5.0	2843	1	US-08-452-654-2
10	154	5.0	2843	1	US-08-452-655B-2
11	154	5.0	2843	1	US-08-452-655B-7
12	154	5.0	2843	2	US-08-370-235A-2
13	154	5.0	2843	3	US-08-450-582-2
14	154	5.0	2843	3	US-08-450-582-7
15	154	5.0	2843	4	US-08-449-731-2
16	154	5.0	2973	2	US-08-821-355A-7
17	154	5.0	2973	2	US-09-003-687A-7
18	154	5.0	2973	3	US-09-136-605-7
19	149.5	4.9	2137	4	US-09-134-001C-4463
20	147.5	4.8	951	4	US-09-125-635-8
21	147.5	4.8	1420	4	US-09-125-635-4
22	139	4.6	725	3	US-09-234-332-12
23	137.5	4.5	1404	4	US-08-801-308-1
24	135	4.4	1911	4	US-09-854-856-64
25	135	4.4	1939	4	US-09-854-856-48
26	135	4.4	1971	4	US-09-854-856-32
27	135	4.4	1999	4	US-09-854-856-16

28	135	4.4	2004	4	US-09-854-856-58	Sequence 58, Appl
29	135	4.4	2032	4	US-09-854-856-42	Sequence 42, Appl
30	135	4.4	2064	4	US-09-854-856-26	Sequence 26, Appl
31	135	4.4	2092	4	US-09-854-856-10	Sequence 10, Appl
32	135	4.4	2157	4	US-09-854-856-52	Sequence 52, Appl
33	135	4.4	2185	4	US-09-854-856-36	Sequence 36, Appl
34	135	4.4	2217	4	US-09-854-856-20	Sequence 20, Appl
35	135	4.4	2245	4	US-09-854-856-4	Sequence 4, Appl
36	134	4.4	3969	3	US-08-061-376-5	Sequence 5, Appl
37	132	4.3	467	4	US-09-134-001C-3235	Sequence 3235, Ap
38	132	4.3	2133	2	US-08-670-707A-37	Sequence 37, Appl
39	132	4.3	2133	3	US-09-037-601-37	Sequence 37, Appl
40	132	4.3	2133	4	US-09-315-179-37	Sequence 37, Appl
41	132	4.3	2133	4	US-09-523-656-30	Sequence 30, Appl
42	131.5	4.3	1402	4	US-09-125-635-12	Sequence 12, Appl
43	130.5	4.3	868	1	US-07-864-004B-6	Sequence 6, Appl
44	130.5	4.3	868	1	US-08-251-937A-6	Sequence 6, Appl
45	130.5	4.3	868	1	US-08-212-133A-3	Sequence 3, Appl

ALIGNMENTS

RESULT 1

US-09-041-886-19
; Sequence 19, Application US/09041886
; Patent No. 6235872
; GENERAL INFORMATION:
; APPLICANT: Bradese, Dale E.
; APPLICANT: Radzede, Sharroz
; TITLE OF INVENTION: Propeptotic Peptides, Dependence
; TITLE OF INVENTION: Polypeptides and Methods of Use
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/041,886
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 2626
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1312 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-041-886-19

Query Match 5.3%; Score 163; DB 3; Length 1312;
Best Local Similarity 19.9%; Pred. No. 9.2e-05;
Matches 120; Conservative 76; Mismatches 260; Indels 146; Gaps 22;

QY	37	ASPLDNKPKQGLSERTGVPKF--KSTPP-----PSLPL-----	71
Db	658	ATPPVARTPSGGTSSVVGVPRLSPKTHRPSPQNSIGNTPSGVPLASPOAGIIPTE	717
QY	72	-----SPPISSPSYFAIPPLGSPAELDSPV-----LLNSSNLPSPTTGAF	114

Db 718 AVAMPIAASPTTASPASNAVTPS---SEAKDRLQDQORQNSPAGNKENIKNETSPSF 774
Qy 115 VAOSFNWSSGGNQIIVKEEDKFSN---FSQTRSGPPASSTATYQSSNVTVQOQPW 171
Db 775 S-----KAENKGISPVVSEHRKQIDDLKFKFNDRLQPSSTSEMDQLLNKREGEKSR 828
Qy 172 SF-----QEAATKQNFSSGKGMKTENSS--SMQSPSEIASVQTNHNSNGFQSDYGNYPQ 225
Db 829 DLKDKIEPSAKDSFIENSSNCTSGSKPNSPSISILS-NTEHKRGPEVTSQGVQTS 887
Qy 226 SOTLSRRSDGYNWRKYQKQVKGSE-----NPRSYKCTYPNCPKTKKVER---S 273
Db 888 SPACKQEKDDKEE-KKDAAEQVRKSTLNPNAKEFNPRSPSQPKPSTTPTSPPQAQSPS 946
Qy 274 LDG--QITEI-----VYKGTIN-HPKPQNTNRNNSNSSSLAIPHNSIR 314
Db 947 MVGHQOQTPVYTPQVCFAPNMMYVPVSPGVQPLYPIMPMPVNOAKTYRAVNNPQOR 1006
Qy 315 TEIPDOSYATH---GSGQMSAATPENSISIGDDDFEQSSOKCKSGGDEYDEPDAGR 371
Db 1007 QDQHOSAMHHPASAAGPPIAATPAYSTQYVAYSQQFPNQLVQHVPHYQSQHPHVS 1066
Qy 372 WKIEGENEGMSAPGSRTRVPRVVVQTTSDIDLDGYNWRKYQKVKGNPNPRSY-- 429
Db 1067 PVIQGNARMWAPP---THAQPLVSSSATQYGAHEQTHAMVACPYPYNKETSPPSYFAI 1123
Qy 430 -----KCTHGPCVPRKHVERASHDLRAVITTYEGKHNHDVPAARGSG-----SHSVNRP 478
Db 1124 STGSLAQQAHPNATLHPHTP-----HPQPSATPTGQQSQHQGSGHP 1165
Qy 479 MPNNASHTNTAATSVLLPVIHQ-----DNSLQNRQSQAPPEGQSPF 522
Db 1166 APSVQHHQAAQALHLASPOQOASAIYHAGLAPTPSPMTPASNTQSPNSPFAAQOTVF 1225
Qy 523 TL 524
Db 1226 TI 1227

RESULT 2
US-09-648-281-2
; Sequence 2, Application US/09648281
; Patent No. 6515197
; GENERAL INFORMATION:
; APPLICANT: Pulst, Stefan M.
; TITLE OF INVENTION: Transgenic Animal Model of
; TITLE OF INVENTION: Neurodegenerative Disease and Methods of Use
; FILE REFERENCE: P-CE 4336
; CURRENT APPLICATION NUMBER: US/09/648,281
; CURRENT FILING DATE: 2000-08-24
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1312
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-648-281-2

Query Match 5.3%; Score 163; DB 4; Length 1312;
Best Local Similarity 19.9%; Pred. No. 9.2e-05;
Matches 120; Conservative 76; Mismatches 260; Indels 146; Gaps 22;

Qy 37 ASPLDNNKPPQGGUSERTGSGVPRF--KSTPP-----PSLPL-----71
Db 658 ATPPVARTSPSGTWSVSGVPRLSKTPRSPRQNSIGNTPSGVPVLASPOAGIIPTE 717
Qy 72 -----SPPPIPSPSYFAIPGLSPAELLDSPV-----LLNSNLSPTTGAF 114
Db 718 AVAMPIAASPTTASPASNAVTPS---SEAKDRLQDQORQNSPAGNKENIKNETSPSF 774
Qy 115 VAOSFNWSSGGNQIIVKEEDKFSN---FSQTRSGPPASSTATYQSSNVTVQOQPW 171

Db 775 S-----KAENKGISPVVSEHRKQIDDLKFKFNDRLQPSSTSEMDQLLNKREGEKSR 828
Qy 172 SF-----QEAATKQNFSSGKGMKTENSS--SMQSPSEIASVQTNHNSNGFQSDYGNYPQ 225
Db 829 DLKDKIEPSAKDSFIENSSNCTSGSKPNSPSISILS-NTEHKRGPEVTSQGVQTS 887
Qy 226 SOTLSRRSDGYNWRKYQKQVKGSE-----NPRSYKCTYPNCPKTKKVER---S 273
Db 888 SPACKQEKDDKEE-KKDAAEQVRKSTLNPNAKEFNPRSPSQPKPSTTPTSPPQAQSPS 946
Qy 274 LDG--QITEI-----VYKGTIN-HPKPQNTNRNNSNSSSLAIPHNSIR 314
Db 947 MVGHQOQTPVYTPQVCFAPNMMYVPVSPGVQPLYPIMPMPVNOAKTYRAVNNPQOR 1006
Qy 315 TEIPDOSYATH---GSGQMSAATPENSISIGDDDFEQSSOKCKSGGDEYDEPDAGR 371
Db 1007 QDQHOSAMHHPASAAGPPIAATPAYSTQYVAYSQQFPNQLVQHVPHYQSQHPHVS 1066
Qy 372 WKIEGENEGMSAPGSRTRVPRVVVQTTSDIDLDGYNWRKYQKVKGNPNPRSY-- 429
Db 1067 PVIQGNARMWAPP---THAQPLVSSSATQYGAHEQTHAMVACPYPYNKETSPPSYFAI 1123
Qy 430 -----KCTHGPCVPRKHVERASHDLRAVITTYEGKHNHDVPAARGSG-----SHSVNRP 478
Db 1124 STGSLAQQAHPNATLHPHTP-----HPQPSATPTGQQSQHQGSGHP 1165
Qy 479 MPNNASHTNTAATSVLLPVIHQ-----DNSLQNRQSQAPPEGQSPF 522
Db 1166 APSVQHHQAAQALHLASPOQOASAIYHAGLAPTPSPMTPASNTQSPNSPFAAQOTVF 1225
Qy 523 TL 524
Db 1226 TI 1227

RESULT 3
US-07-741-940-7
; Sequence 7, Application US/07741940
; Patent No. 5352775
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESSES:
; ADDRESS: Banner, Birch, McKie & Beckett
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/741,940
; FILING DATE: 19920109
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.035574

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-508-9100
 TELEFAX: 202-508-9299
 INFORMATION FOR SEQ ID NO: 7:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2842 amino acids
 TYPE: AMINO ACID
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 ORIGINAL SOURCE:
 ORGANISM: Homo sapiens
 IMMEDIATE SOURCE:
 CLONE: APC
 US-07-741-940-7

Query Match 5.0%; Score 154; DB 1; Length 2842;
 Best Local Similarity 20.6%; Pred. No. 0.0017;
 Matches 124; Conservative 66; Mismatches 235; Indels 176; Gaps 25;

QY 30 TSPSLLASPLDNNKP--PQG-----GLSERTGSGVPKFKSTPPPSLPLSPSPSS 80
 Db 2206 SEISGQMKQLOANMPISGRMTIHIPGVNRSSSTSPVSKGPPKTPASKSPSEGQT 2265
 QY 81 YFAIPPGGLSP---AELLDSPVLLNSSNI-----LPSPTTGAFVAQSF 119
 Db 2266 ATTSRGAKEPVKSEL--SPVARQTSQIGSSKAPSRGSRDSTPSRPAQQLSRPIQSP 2323
 QY 120 NWKSSGCGNOQIVKEEDKFSNFSFQTRSGPPASSTATYQSSNVTVQTO-QPMSFQEA TK 178
 Db 2324 GRNSISPRNGISPPNKLKLS---QLPRTSSPSTASTKSSGSGKMSYTPSGRQMSQQNLTK 2379
 QY 179 QDNFSSCKGMKMTENSSMSQSPSEIASVQTHNSNGFQSDYGNYPQSQTLRSRSDGYN 238
 Db 2380 QTGLS-----KNASSIP--RSESASKGLNQMN-----GNGANKKVELSRMSS----- 2420
 QY 239 WRKYGQKQVKGSENPVSYYKCTY-----PNCPTKKKVERSLDQITETVYKGTNHPKPQN 294
 Db 2421 -TKSSGESDRSERPVLVROSTFIKEAPSTLRKLEESAS-----FESLSPSRPAS 2472
 QY 295 TRNNSNSSSLAIPHNSIRTEIPDQSYATHGSGQMS-AATPENSISIGDDDDFQSSQ 353
 Db 2473 PTRSQAQTPVLS-----PSLPDMSLSTHSSVQAGGWRKLPNLSPTI----- 2514
 QY 354 KCKSGGDEYDEDAKR-----WKIGENEGMAGSRTVREP 392
 Db 2515 -----EYNDGRP-AKHDIARSHSESRPLPINRGTWKREKSHSSLPVST----- 2562
 QY 393 RVVVQTTSDIDLDGVRWKYQKVVKGPNPNRVSYYKCTHPGCPVRKHVERASHDLRAV 452
 Db 2563 -----WRTGSSSSILSASSESEKAKSED---EKHV-----NSI 2594
 QY 453 ITTYEGKHNDVPAARG-----SGSHSVNRPMPNNAHNTNTAATSVLLPVIHQ 502
 Db 2595 SGTKQSKENQ--VSAKGTWRKIKENEPSTNSTQTVSSGATNGAESKTLIYQMAPAVK 2652
 QY 503 S-----DNSLQNRQAPPEQSPFTLEMLQSPGFSFGNPMQSYVNVQQLSDN 554
 Db 2653 TEDVWVRIEDCPINNPRSGRSPTGNTTPFVIDSV-----SEKANPNIKSKDNQAKQN 2704
 QY 555 V 555
 Db 2705 V 2705

RESULT 4

US-08-289-548A-7
 Sequence 7, Application US/08289548A
 Patent No. 5648212
 GENERAL INFORMATION:
 APPLICANT: ALBERTSEN, HANS
 APPLICANT: ANAND, RAKESH
 APPLICANT: CARLSON, MARY

APPLICANT: GRODEN, JOANNA
 APPLICANT: HEDGE, PHILIP J.
 APPLICANT: JOSLYN, GEOFF
 APPLICANT: KINZLER, KENNETH
 APPLICANT: MARKHAM, ALEXANDER F.
 APPLICANT: NAKAMURA, YUSUKE
 APPLICANT: THLIVERIS, ANDREW
 TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
 TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
 NUMBER OF SEQUENCES: 102
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Banner & Allegretti, LTD
 STREET: 1001 G Street, NW
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20001-4598
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/289,548A
 FILING DATE: 12-AUG-1994
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Kagan, Sarah A.
 REGISTRATION NUMBER: 32,141
 REFERENCE/DOCKET NUMBER: 1107.46943
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-508-9100
 TELEFAX: 202-508-9299
 INFORMATION FOR SEQ ID NO: 7:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2842 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 ORIGINAL SOURCE:
 ORGANISM: Homo sapiens
 IMMEDIATE SOURCE:
 CLONE: APC
 US-08-289-548A-7

Query Match 5.0%; Score 154; DB 1; Length 2842;
 Best Local Similarity 20.6%; Pred. No. 0.0017;
 Matches 124; Conservative 66; Mismatches 235; Indels 176; Gaps 25;

QY 30 TSPSLLASPLDNNKP--PQG-----GLSERTGSGVPKFKSTPPPSLPLSPSPSS 80
 Db 2206 SEISGQMKQLOANMPISGRMTIHIPGVNRSSSTSPVSKGPPKTPASKSPSEGQT 2265
 QY 81 YFAIPPGGLSP---AELLDSPVLLNSSNI-----LPSPTTGAFVAQSF 119
 Db 2266 ATTSRGAKEPVKSEL--SPVARQTSQIGSSKAPSRGSRDSTPSRPAQQLSRPIQSP 2323
 QY 120 NWKSSGCGNOQIVKEEDKFSNFSFQTRSGPPASSTATYQSSNVTVQTO-QPMSFQEA TK 178
 Db 2324 GRNSISPRNGISPPNKLKLS---QLPRTSSPSTASTKSSGSGKMSYTPSGRQMSQQNLTK 2379
 QY 179 QDNFSSCKGMKMTENSSMSQSPSEIASVQTHNSNGFQSDYGNYPQSQTLRSRSDGYN 238
 Db 2380 QTGLS-----KNASSIP--RSESASKGLNQMN-----GNGANKKVELSRMSS----- 2420
 QY 239 WRKYGQKQVKGSENPVSYYKCTY-----PNCPTKKKVERSLDQITETVYKGTNHPKPQN 294
 Db 2421 -TKSSGESDRSERPVLVROSTFIKEAPSTLRKLEESAS-----FESLSPSRPAS 2472
 QY 295 TRNNSNSSSLAIPHNSIRTEIPDQSYATHGSGQMS-AATPENSISIGDDDDFQSSQ 353
 Db 2473 PTRSQAQTPVLS-----PSLPDMSLSTHSSVQAGGWRKLPNLSPTI----- 2514

```
QY 354 KCKSGGDEYDEBPAKR-----WKIEGENEGMSAPGSRVREP 392
Db 2515 -----EYNDGRP-AKRHDIAHSHSPRLPINRSGTWKRKSHKSSLPVST----- 2562
QY 393 RVVVQTTSDIDLDGGRWKYQGVKVGKPNPRSYKCTHPCGPKVKKHVERASHDLRAV 452
Db 2563 -----WRTGSSSSILSASSESEKAKSED---EKHV-----NSI 2594
QY 453 ITTYEGKHNHVDVPAARG-----SGSHSVNRPMPNPNASHTNTAATSVLLPVIHQ 502
Db 2595 SGTQKSKENQ--VSAKGTWRKIKENEFSTNTSTQTVSSGATNGAESKTLIYQMAPAVSK 2652
QY 503 S-----DNSLQNRQSQAPPEGQSPFTLEMLQSPGFGFQGNPMQSYVNVQQQLSDN 554
Db 2653 TEDVWVRIEDCPINNPRSGRSGPTGNTPPVIDSV-----SEKANPNIKDKDNQAKQN 2704
QY 555 V 555
Db 2705 V 2705

RESULT 5
US-08-452-654-7
; Sequence 7, Application US/08452654
; Patent No. 5691454
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckett
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,654
; FILING DATE: 25-MAY-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/741,940
; FILING DATE: 08-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.035574
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2842 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
```

```
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: APC
US-08-452-654-7

Query Match 5.0%; Score 154; DB 1; Length 2842;
Best Local Similarity 20.6%; Pred. No. 0.0017;
Matches 124; Conservative 66; Mismatch 235; Indels 176; Gaps 25;

QY 30 TFSFDLLASPLDNKP--PQG-----GLSERTSGVPPKPKSTPPSPPLSPSPSPSS 80
Db 2206 SEISGQMKQFLOANPMSISRGRTMIHPCVRNSSSTSPVSKGPKLTKTPASKSPSEGT 2265
QY 81 YFAIPPLGLSP--AEILDSPVLLNSSNI-----LPSPTTGAFAVQSF 119
Db 2266 ATTSRPAKPSVKSEL--SPVARQTSQIGSSKAPSRSGSRDSTSRPAQPLSRPIQSP 2323
QY 120 NWKSSGNGNQIIVKEEDKSFNFQTRSGPPASSTATYQSSNVTVQTO-QPMSFOEATK 178
Db 2324 GRNSISPGRNGISPPNKLKLS---QLPRTSSPSTASTKSSGSGKMSYTSPPGRQMSQNLTK 2379
QY 179 QDNFSSGKGMKMTENSSSMQSPFSPETASVQTHNSGFDQSDYGNYPPOSQTLRRSDDGYN 238
Db 2380 QTGLS-----KNASSIP--RSESASKGLNQMN-----GNGANKKVELSRMSS--- 2420
QY 239 WRKYQGVKQVKGSENPSPSYKCTY---PNCPTKKKVERSLEDGQITEIVYKGTNHPKPQN 294
Db 2421 -TKSGSESDRSERPVLVRQSTFIKEAPSTLRKLEESAS-----FESLSPSRPAS 2472
QY 295 TRRNSNSSSLAIPHNSIRTEIPDOSYATHSGSQMDS-AATPENSSISIGDDDFQSSQ 353
Db 2473 PIRSQATPVLS-----PSLPDMSLSTHSSVQAGWRKLPNLSPTI----- 2514
QY 354 KCKSGGDEYDEBPAKR-----WKIEGENEGMSAPGSRVREP 392
Db 2515 -----EYNDGRP-AKRHDIAHSHSPRLPINRSGTWKRKSHKSSLPVST----- 2562
QY 393 RVVVQTTSDIDLDGGRWKYQGVKVGKPNPRSYKCTHPCGPKVKKHVERASHDLRAV 452
Db 2563 -----WRTGSSSSILSASSESEKAKSED---EKHV-----NSI 2594
QY 453 ITTYEGKHNHVDVPAARG-----SGSHSVNRPMPNPNASHTNTAATSVLLPVIHQ 502
Db 2595 SGTQKSKENQ--VSAKGTWRKIKENEFSTNTSTQTVSSGATNGAESKTLIYQMAPAVSK 2652
QY 503 S-----DNSLQNRQSQAPPEGQSPFTLEMLQSPGFGFQGNPMQSYVNVQQQLSDN 554
Db 2653 TEDVWVRIEDCPINNPRSGRSGPTGNTPPVIDSV-----SEKANPNIKDKDNQAKQN 2704
QY 555 V 555
Db 2705 V 2705

RESULT 6
US-08-449-731-7
; Sequence 7, Application US/08449731
; Patent No. 6413727
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 102
```

;;
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Banner & Allegretti, LTD
;; STREET: 1001 G Street, NW
;; CITY: Washington
;; STATE: D.C.
;; COUNTRY: USA
;; ZIP: 20001-4598
;;
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent In Release #1.0, Version #1.25
;;
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/449,731
;; FILING DATE: 25-May-1995
;; CLASSIFICATION: <Unknown>
;;
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: 08/289,548
;; FILING DATE: 12-AUG-1994
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Kagan, Sarah A.
;; REGISTRATION NUMBER: 32,141
;; REFERENCE/DOCKET NUMBER: 1107.46943
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 202-508-9100
;; TELEFAX: 202-508-9299
;;
;; INFORMATION FOR SEQ ID NO: 7:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 2842 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; ORIGINAL SOURCE:
;; ORGANISM: Homo sapiens
;; CLONE: APC
;; IMMEDIATE SOURCE:
;; SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-08-449-731-7

Query Match 5.0%; Score 154; DB 4; Length 2842;
Best Local Similarity 20.6%; Pred. No. 0.0017;
Matches 124; Conservative 66; Mismatches 235; Indels 176; Gaps 25;

QY 30 TSPSDDLASPLDNNKP--PGQ-----GLSERTGSGVPRKFKSTPPPSLPLSPPIPSPS 80
DB 2206 SEISGQMKQLOANMPSISGRMTIHIPGVNRSSSTSPVSKGPPKLTASKSPSEGT 2265

QY 81 YFAIPPGLSPP---AELDSPVLLNMMI-----LPSPTTGAFVAQSF 119
DB 2266 ATTSRCAKPSVKSEL--SPVARQTSQIGSSKAPSRSGSRDSTPSRPAQQLSRPIQSP 2323

QY 120 NWKSSGCGNQIVKEEDKSNFSFQTRGPPASSTATYQSSNVTVQTO-QPWSFQBATK 178
DB 2324 GRNSISPRNGISPPNKLKLS---QLPRTSSPSTASTKSSGSGKMSYTSPPGRQMSQQNLTK 2379

QY 179 QDNFSSGKGWMTNSSSMQSFSEIASVQTNHNGFQSDYGNVPPQSTLSRSDGYN 238
DB 2380 QTGLS-----KNASSIP--RSESASKGLNQMN-----GNGANKKVELSRMS-- 2420

QY 239 WRKYGQKQVKGSENPRIYKCTY-----PNCPTKKKVERSLDQGITETIVYKGTNNHPQN 294
DB 2421 -TKSSGESDESERPLVROSTFFKEAPSTLRKLEESAS-----PESLSPSRPAS 2472

QY 295 TRNNSNSSSLAIPHNSIRTEIPDQSYATHGSCQMS--AATPNSISIGDDDFEQSSQ 353
DB 2473 PTRSQATPVLK-----PSLDMSLSTHSSVQAGGWRKLPNLSPTI----- 2514

QY 354 KCKSGGDEYDEPDAR-----WKIEGECMGAPGSRVREP 392
DB 2515 -----EYNDGRP-AKRHDIAHSHSESRLPINRSGTWKREHSHSSSLPRVST---- 2562

QY 393 RVVVQTTSDIDLDGGRWKYQKVVKNPNPRSYKCTHPGCPVKHVERASHDLRAV 452

Db 2563 -----WRRTGSSSSILSASSESEKAKSD-----EKHV-----NSI 2594
QY 453 ITTYECKHNHDVPAARG-----SGSHSVNRPMNNASNHTTAATSVRELLPVIHQ 502
Db 2595 SGTQSKENQ--VSAKGTWRKIKENEFSPINSTSTQTVSSGATNGAESKTLIYQMAVSK 2652
QY 503 S-----DNSLQNRQSQAPPEGQSPFTTLEMLQSPGSGFSGFGNPMOSYVNNQQQLSDN 554
Db 2653 TEDVWVRIEDCPINNPRSGRSPGTGTPPVIDSV-----SEKANPNIKDSKDNQAKQN 2704
QY 555 V 555
Db 2705 V 2705

RESULT 7
US-07-741-940-2
; Sequence 2, Application US/07741940
; Patent No. 5352775
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MAKHAM, ALEXANDER F.
; APPLICANT: MAKAMURA, YUSUKE
; APPLICANT: THILVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckett
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/741,940
; FILING DATE: 19920109
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.035574
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2843 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-741-940-2

Query Match 5.0%; Score 154; DB 1; Length 2843;
Best Local Similarity 20.6%; Pred. No. 0.0017;
Matches 124; Conservative 66; Mismatches 235; Indels 176; Gaps 25;

QY 30 TSPSDDLASPLDNNKP--PGQ-----GLSERTGSGVPRKFKSTPPPSLPLSPPIPSPS 80
DB 2207 SEISGQMKQLOANMPSISGRMTIHIPGVNRSSSTSPVSKGPPKLTASKSPSEGT 2266

```

QY 81 YFAIPPLGLSP---AELLDSPVLLNSSNI-----LPSPTTGAFVAQSF 119
Db 2267 ATTSPRGAKSVKSEL--SPVARQTSIGSSKAPSGSRDSTPSRPAQQLSRPIQSP 2324
QY 120 NWKSSGNGNOIQVEEDKSFNSFQTRSGPPASSTATYOSSNVTVQTO-QPWSFQEA TK 178
Db 2325 GRNSISPGRNGISPPNKLKLS---QLPRTSSPSTASTKSSGSGKMSYTSPPGRQMSQQNLTK 2380
QY 179 QDNFSSGCKGMKWTENSSMSQSFPEIASVQTNHNSGFSQSDYGNVPPQSQTLSRRSDGYN 238
Db 2381 QTGLS-----KNASSIP--RSESASGKLNQNN-----GNGANKVELSRMSS----- 2421
QY 239 WRKYGQKVGKSENPRSYKCTY-----PNCPTKKKVERSLEDGQITTEIVYKGTNHPKPN 294
Db 2422 -TKSSGESDRSERPVLVRQSTFIKEAPSTLRRKLEESAS-----PESLSPSRPAS 2473
QY 295 TRNNSNSSSLAIPHNSIRTEIPDQSYATHSGQWDS-AATPENSISIGDDDDFEQSSQ 353
Db 2474 PTRSQATPVLS-----PSLPDMSLSTHSSVQAGGWRKLPNLSPTI----- 2515
QY 354 KCKSGGDEYDEDEPAKR-----WKIEGENEGMSAPGSRVTREP 392
Db 2516 -----EYNDGRP-AKRHDIAKSHSESPSRLPINRSGTWKREHSHSSLPVST----- 2563
QY 393 RVVVQTTSDIDLDGVRWKYQKVVKGPNPRSYKCTHPGCPVKRKHVERASHDLRAV 452
Db 2564 -----WRTGSSSILSASSESEKAKSED-----EKHV-----NSI 2595
QY 453 ITTYEGKHNHDVPAARG-----SGSHSVNRMPNPNNASHTNTAATSVRLLPVHQ 502
Db 2596 SGTQSKENQ--VSAKGTWRKIKENBFPSTNSTQTVSSGATNGAESKTLIYQMAPAVSK 2653
QY 503 S-----DNSLONQSOAPPEGQSPFTLEMLQSPGSGFGNGPMQSVNQOQLSDN 554
Db 2654 TEDVWVRIEDCPINNPRSGRSPGTNTPPVIDSV-----SEKANPNIKSDKNQAKQN 2705
QY 555 V 555
Db 2706 V 2706

```

RESULT 8

```

US-08-289-548A-2
; Sequence 2, Application US/08289548A
; Patent No. 5648212
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Allegretti, LTD
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/289,548A

```

```

; FILING DATE: 12-AUG-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.46943
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2843 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-289-548A-2

Query Match 5.0%; Score 154; DB 1; Length 2843;
Best Local Similarity 20.6%; Pred. No. 0.0017;
Matches 124; Conservative 66; Mismatches 235; Indels 176; Gaps 25;

QY 30 TSFSDLLASPLDNKP--PQG-----GLSERTSGVPKFKSTPPSLPLSPPIPSPS 80
Db 2207 SEISGQMKQFLOANPMSIRGRMTIHIPGVRNSSSTSPVSKGKPLKTPAKSPSEGT 2266
QY 81 YFAIPPLGLSP---AELLDSPVLLNSSNI-----LPSPTTGAFVAQSF 119
Db 2267 ATTSPRGAKSVKSEL--SPVARQTSIGSSKAPSGSRDSTPSRPAQQLSRPIQSP 2324
QY 120 NWKSSGNGNOIQVEEDKSFNSFQTRSGPPASSTATYOSSNVTVQTO-QPWSFQEA TK 178
Db 2325 GRNSISPGRNGISPPNKLKLS---QLPRTSSPSTASTKSSGSGKMSYTSPPGRQMSQQNLTK 2380
QY 179 QDNFSSGCKGMKWTENSSMSQSFPEIASVQTNHNSGFSQSDYGNVPPQSQTLSRRSDGYN 238
Db 2381 QTGLS-----KNASSIP--RSESASGKLNQNN-----GNGANKVELSRMSS----- 2421
QY 239 WRKYGQKVGKSENPRSYKCTY-----PNCPTKKKVERSLEDGQITTEIVYKGTNHPKPN 294
Db 2422 -TKSSGESDRSERPVLVRQSTFIKEAPSTLRRKLEESAS-----PESLSPSRPAS 2473
QY 295 TRNNSNSSSLAIPHNSIRTEIPDQSYATHSGQWDS-AATPENSISIGDDDDFEQSSQ 353
Db 2474 PTRSQATPVLS-----PSLPDMSLSTHSSVQAGGWRKLPNLSPTI----- 2515
QY 354 KCKSGGDEYDEDEPAKR-----WKIEGENEGMSAPGSRVTREP 392
Db 2516 -----EYNDGRP-AKRHDIAKSHSESPSRLPINRSGTWKREHSHSSLPVST----- 2563
QY 393 RVVVQTTSDIDLDGVRWKYQKVVKGPNPRSYKCTHPGCPVKRKHVERASHDLRAV 452
Db 2564 -----WRTGSSSILSASSESEKAKSED-----EKHV-----NSI 2595
QY 453 ITTYEGKHNHDVPAARG-----SGSHSVNRMPNPNNASHTNTAATSVRLLPVHQ 502
Db 2596 SGTQSKENQ--VSAKGTWRKIKENBFPSTNSTQTVSSGATNGAESKTLIYQMAPAVSK 2653
QY 503 S-----DNSLONQSOAPPEGQSPFTLEMLQSPGSGFGNGPMQSVNQOQLSDN 554
Db 2654 TEDVWVRIEDCPINNPRSGRSPGTNTPPVIDSV-----SEKANPNIKSDKNQAKQN 2705
QY 555 V 555
Db 2706 V 2706

RESULT 9
US-08-452-654-2
; Sequence 2, Application US/08452654
; Patent No. 5691454
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH

```

APPLICANT: CARLSON, MARY
APPLICANT: GRODEN, JOANNA
APPLICANT: HEDGE, PHILIP J.
APPLICANT: JOSLYN, GEOFF
APPLICANT: KINZLER, KENNETH
APPLICANT: MARKHAM, ALEXANDER F.
APPLICANT: NAKAMURA, YUSUKE
APPLICANT: THLIVERIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
NUMBER OF SEQUENCES: 94
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner, Birch, McKie & Beckett
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/452,654
FILING DATE: 25-MAY-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/741,940
FILING DATE: 08-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.035574
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9299
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2843 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-452-654-2

Query Match 5.0%; Score 154; DB 1; Length 2843;
Best Local Similarity 20.8%; Pred. No. 0.0017;
Matches 124; Conservative 66; Mismatches 235; Indels 176; Gaps 25;

QY 30 TFSFDLLASPLDNNKP--POG-----GLSERTGSGVPKFKSTPPPLSLPPSPISPS 80
DB 2207 SEISGQMKQPLQANMPSISGRFTMIHPGVNRSSSTSPVSKGPKLPKTPASKSPSEGQT 2266
QY 81 YFAIPPLGLP---AELLDGVLNLSNI-----LPSFTTGAFVAQSF 119
DB 2267 ATTSRGAQPSVKSEL--SPVARQTSIGGSSKAPSRSGRSDTSPRPAQQLSRRIQSP 2324
QY 120 NKKSSGGNQIIVKEDKFSNPSFQTRSGPPASSTATVQSSNVTQTO-QPWSFQEA TK 178
DB 2325 GRNISPGRNGISPPNKLKLS-----QLPRTSPSTASTKSGSGKMSVTSFGRCMSQQLTK 2380
QY 179 QDNFSSGKGMMKTESSSMQSPFELASVQTNHNGFQSDYGNYPQSQOTLSRSRSDGYN 238
DB 2381 QTGLS-----KNASTIP--RSEASKGLNOMN-----NGANKVELSRMSS----- 2421
QY 239 WRKYGQKQVGSNPRSYKCTY-----PNCPTKKKVERSLDQITBIVYKGTNNHPKQPN 294
DB 2422 -TKSSGESDRSRPVLVRQSTTIKAAPSLTRKLEESAS-----FESLSPSRSPAS 2473
QY 295 TRNNSNSSSLAIPHNSINTETPDOSYATHGSGQMDS-AATPENSISIGDDDFQSQSQ 353
DB 2474 PTRSQATPVLS-----PSLPDMSLSTHSSVQAGWRKLPPLNLSPTI----- 2515

RESULT 10

US-08-452-655B-2
; Sequence 2, Application US/08452655B
; Patent No. 5783666
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Witcoff, Ltd.
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,655B
; FILING DATE: 25-MAY-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/289,548
; FILING DATE: 12-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/741,940
; FILING DATE: 08-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.49964
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2843 amino acids
; TYPE: amino acid
; TOPOLOGY: linear

QY 354 KCKSGGDEYDEPDPAKR-----WKTEGENEGMSAFGSRVTRP 392
DB 2516 -----EYNDGRP-AKRHDIAARSHSESPRLPINRSGTWKREHSGHSSSLPRVST----- 2563
QY 393 RVVVQTTSDIDLDGGRKRYGQKVVKGNPNRPSYKCTHFGCPCVRKHVERASHDLRAV 452
DB 2564 -----WRRTGSSSSILSASESESEKAKSED-----ERHV-----NSI 2595
QY 453 ITTYEGKHNHDVPAARG-----SGSHSVNRPMPNNAHNTHTNTATSVRLLPVIHQ 502
DB 2596 SGTQSKENQ--VSAGTWRKIKENEFSPNTSTQTVSSGATGAESKTLTIQMAPAVSK 2653
QY 503 S-----DNSLQNRQAPPEGQSPFTLEMLQSPGSPGFGNPMOSYVNOQLSDN 554
DB 2654 TEDVWVRIEDCPINNPRSGRSPGTGNTPPVIDSV-----SEKANPNIKDSKDNQAKQN 2705
QY 555 V 555
DB 2706 V 2706

MOLECULE TYPE: protein

US-08-452-655B-2

Query Match 5.0%; Score 154; DB 1; Length 2843;

Best Local Similarity 20.6%; Pred. No. 0.0017; Matches 124; Conservative 66; Mismatches 235; Indels 176; Gaps 25;

```

QY 30 TSFSDLLASPLDNKP--PG-----GLSERTGCVKFKSTPPPSLPLSPSPSPS 80
DB 2207 SEISGQKQLOANMPSISGRMTMIHIPGVNRSSSTSPVSKGPKLTPASKSPSEGQT 2266
QY 81 YFAIPPLGSP---AELDSPLVLLNSNI-----LPSPTTCAFAVQSF 119
DB 2267 ATTSRGAKEPSVKSEL--SPVARQTSQIGSSKAPSRSGSRDSTPSRPAQQLSRFIQSP 2324
QY 120 NWKSSSGGNQIQVKEEDKSFNFSTQTRSGPPASSTATYQSSNVTVQTO-QPMSFQEA TK 178
DB 2325 GRNSISPGRNGISPPNKL-----QLPRTSSPSTASTKSSGSGKMSYTSRQMSQNLTK 2380
QY 179 QDNFSSGKGMKMTENSSSMQSFPEIASVQTNHNGFQSDYGNYPQSQTLRRSDGYN 238
DB 2381 QTGLS-----KNASSIP--RSESASKGLNOMN-----GNANKKVELSRMSS-- 2421
QY 239 WRKYGQKQVKGSENPRSYVKCTY----PNCPTKKKVERSILDGOITEIVYKGTNHPKPQN 294
DB 2422 -TKSSGESDRSRPVLVRQSTFIKEAPSTLRKLEESAS-----FESLSPSRPAS 2473
QY 295 TRNNSNSSSLAIPHNSIRTEIPDOSYATHGSGQMDS-AATPENSISIGDDDDFEQSSQ 353
DB 2474 PTRSQAQTPVLS-----PSLPDMSLTHSSVQAGGWRKLPNLSPTI----- 2515
QY 354 KCKSGGDEYDEPDADR-----WKIEGENEGMSAPGSRVREP 392
DB 2516 -----EYNDGRP-AKRHDIAHSHSESRPLINRSGTWKREHSHSSLPVST----- 2563
QY 393 RVVVQTTSDIDILDDGVRWKYQKVKGNPNPRSYKCTHPGCPVRKHVERASHDLRAV 452
DB 2564 -----WRTGSSSILSASSESEKAKSED---EKHV-----NSI 2595
QY 453 ITTYEGKHNDVPAARG-----SGSHSVNRPMPNNAHNTNTAATSVLLPVIHQ 502
DB 2596 SGTQKSKENGQ--VSAKGTWRKIKENEPSTNSTSQTVSSGATGAESKTLIYQMAPAVSK 2653
QY 503 S-----NSLQNRQSQAPPECQSPFTLEMLQSPGSGFGNGNPMQSVNQQLSDN 554
DB 2654 TEDVWVRIEDCPINPRSGRSPTNTPPVIDSV-----SEKANPNIKDKDNQAKQN 2705
QY 555 V 555
DB 2706 V 2706

```

RESULT 11

US-08-452-655B-7
Sequence 7, Application US/08452655B

Patent No. 5783666

GENERAL INFORMATION:

APPLICANT: ALBERTSEN, HANS

APPLICANT: ANAND, RAKESH

APPLICANT: CARLSON, MARY

APPLICANT: GRODEN, JOANNA

APPLICANT: HEDGE, PHILIP J.

APPLICANT: JOSLYN, GEOFF

APPLICANT: KINZLER, KENNETH

APPLICANT: MARKHAM, ALEXANDER F.

APPLICANT: NAKAMURA, YUSUKE

APPLICANT: THLIVERTS, ANDREW

TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC

TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS

NUMBER OF SEQUENCES: 102

CORRESPONDENCE ADDRESS:

ADDRESSEE: Banner & Witcoff, Ltd.

STREET: 1001 G Street, NW

```

; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,655B
; FILING DATE: 25-MAY-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/289,548
; FILING DATE: 12-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/741,940
; FILING DATE: 08-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.49964
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2843 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEetical: YES
; ANTI-SENSE: NO
; US-08-452-655B-7

```

Query Match 5.0%; Score 154; DB 1; Length 2843;

Best Local Similarity 20.6%; Pred. No. 0.0017;

Matches 124; Conservative 66; Mismatches 235; Indels 176; Gaps 25;

```

QY 30 TSFSDLLASPLDNKP--PG-----GLSERTGCVKFKSTPPPSLPLSPSPSPS 80
DB 2207 SEISGQKQLOANMPSISGRMTMIHIPGVNRSSSTSPVSKGPKLTPASKSPSEGQT 2266
QY 81 YFAIPPLGSP---AELDSPLVLLNSNI-----LPSPTTCAFAVQSF 119
DB 2267 ATTSRGAKEPSVKSEL--SPVARQTSQIGSSKAPSRSGSRDSTPSRPAQQLSRFIQSP 2324
QY 120 NWKSSSGGNQIQVKEEDKSFNFSTQTRSGPPASSTATYQSSNVTVQTO-QPMSFQEA TK 178
DB 2325 GRNSISPGRNGISPPNKL-----QLPRTSSPSTASTKSSGSGKMSYTSRQMSQNLTK 2380
QY 179 QDNFSSGKGMKMTENSSSMQSFPEIASVQTNHNGFQSDYGNYPQSQTLRRSDGYN 238
DB 2381 QTGLS-----KNASSIP--RSESASKGLNOMN-----GNANKKVELSRMSS-- 2421
QY 239 WRKYGQKQVKGSENPRSYVKCTY----PNCPTKKKVERSILDGOITEIVYKGTNHPKPQN 294
DB 2422 -TKSSGESDRSRPVLVRQSTFIKEAPSTLRKLEESAS-----FESLSPSRPAS 2473
QY 295 TRNNSNSSSLAIPHNSIRTEIPDOSYATHGSGQMDS-AATPENSISIGDDDDFEQSSQ 353
DB 2474 PTRSQAQTPVLS-----PSLPDMSLTHSSVQAGGWRKLPNLSPTI----- 2515
QY 354 KCKSGGDEYDEPDADR-----WKIEGENEGMSAPGSRVREP 392
DB 2516 -----EYNDGRP-AKRHDIAHSHSESRPLINRSGTWKREHSHSSLPVST----- 2563
QY 393 RVVVQTTSDIDILDDGVRWKYQKVKGNPNPRSYKCTHPGCPVRKHVERASHDLRAV 452
DB 2564 -----WRTGSSSILSASSESEKAKSED---EKHV-----NSI 2595

```


QY 453 ITTYEGKHNDVPAARG-----SGSHSVNRPMPNNAHNTNTAATSVRLLPVHQ 502
Db 2596 SGTQSKENQ--VSAKGTWRKIKENEFSPNTNSTQTVSSGATNGAESKTLIYQMAPAVSK 2653
QY 503 S-----DNSLONORSQAPPEQSPPTLEMLQSPGSGFGNPMQSVYVNOQQLSDN 554
Db 2654 TEDVWVRIEDCPINNPRSGRSPGTNTPPVIDSV-----SEKANPNIKSDKNQAKQN 2705
QY 555 V 555
Db 2706 V 2706

RESULT 12
US-08-370-235A-2
; Sequence 2, Application US/08370235A
; Patent No. 5910418
; GENERAL INFORMATION:
; APPLICANT: VOGELSTEIN, BERT
; APPLICANT: KINZLER, KENNETH W.
; APPLICANT: HILL, DAVID E.
; APPLICANT: JOHNSON, KAREN A.
; TITLE OF INVENTION: ANTIBODIES AND ASSAYS FOR DERMING
; TITLE OF INVENTION: MUTATIONS IN THE APC GENE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BANNER & WITCOFF, LTD.
; STREET: 1001 G STREET, N.W.
; CITY: WASHINGTON
; STATE: DC
; COUNTRY: US
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/370,235A
; FILING DATE: 01-JAN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: KAGAN, SARAH A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 01107.48688
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 508 9100
; TELEFAX: 202 508 9299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2843 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-370-235A-2

Query Match 5.0%; Score 154; DB 2; Length 2843;
Best Local Similarity 20.6%; Pred. No. 0.0017;
Matches 124; Conservative 66; Mismatches 235; Indels 176; Gaps 25;

QY 30 TSFSDLLASPLDNNKP--PQG-----GLSERTGCVKFKSTPPSLPLSPPIPSPS 80
Db 2207 SEISQMKQPLQANMPSISRGRTMIHIPGVNRSSSTSPVSKGPPPLKTPASKSPBEGQT 2266
QY 81 YFAIPGGLSP---AELLDSPLVLLNSNI-----LPSPTTGAFVAQSF 119
Db 2267 ATTSFPGAKPSVKSEL--SPVARQTSQIGGSKAPSRSGSRDSTPPSPAQQPLSRPIQSP 2324
QY 120 NWKSSSGNQIQVKEEDKSFNSFQTRSGPPASSTATYOSNVTVQTO-QPKWSFQBATK 178
Db 2325 GRNSISFGRNGISPPNKLK-----QLPRTSPFSTASTKSSGKRWSTYSPGRQWSQQLTK 2380
QY 179 QDNFSSGKGMKTNSSSMQSFSPFIASVQTNHNSGFSQDYGNYPPQSQTLRRSRDDGYN 238

Db 2381 QTGLS-----KNASSIP--RSESASKGLNQMN--GNANKKVELSRMSS-- 2421
QY 239 WRKYGQKQVKGSENPRSYKCTV---PNCPTKKKVERSILDGOITEIVYKGTNHPKPN 294
Db 2422 -TKSSGESDRSRPVLVQSTFIKEAPSTLRKLEBAS-----FESLPSRRPAS 2473
QY 295 TRNNSNSSSLAIPHSNIRTEIPDOSYATHSGQMDS-AATPENSISIGDDDDFEQSSQ 353
Db 2474 PTRSQATPVL-----PSLPDMSLSLTHSSVQAGWRKLPPLNLSPTI----- 2515
QY 354 KCKSGGDEYDEPDADR-----WKIEGENEGHSAFGSRTVREP 392
Db 2516 -----EYNDGRP-AKRHDIAHSHSESPRLPINRSGTWKREHSHSSSLPRVST----- 2563
QY 393 RVVVQTSDIDLDGVRWKYQKVVKGPNPRSYKCTHPGCPVKRKHVERASHDLRAV 452
Db 2564 -----WRTGSSSSILSASESESEKAKSED---EKHV-----NSI 2595
QY 453 ITTYEGKHNDVPAARG-----SGSHSVNRPMPNNAHNTNTAATSVRLLPVHQ 502
Db 2596 SGTQSKENQ--VSAKGTWRKIKENEFSPNTNSTQTVSSGATNGAESKTLIYQMAPAVSK 2653
QY 503 S-----DNSLONORSQAPPEQSPPTLEMLQSPGSGFGNPMQSVYVNOQQLSDN 554
Db 2654 TEDVWVRIEDCPINNPRSGRSPGTNTPPVIDSV-----SEKANPNIKSDKNQAKQN 2705
QY 555 V 555
Db 2706 V 2706

RESULT 13
US-08-450-582-2
; Sequence 2, Application US/08450582
; Patent No. 6114124
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Witcoff, Ltd.
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/450,582
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/452,655
; FILING DATE: 25-MAY-1995
; APPLICATION NUMBER: US 08/289,548
; FILING DATE: 12-AUG-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/741,940

;; FILING DATE: 08-AUG-1991
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Kagan, Sarah A. 32,141
;; REGISTRATION NUMBER: 32,141
;; REFERENCE/DOCKET NUMBER: 1107.49964
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 202-508-9100
;; TELEFAX: 202-508-9299
;; INFORMATION FOR SEQ ID NO: 2:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 2843 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; US-08-450-582-2

Query Match 5.0%; Score 154; DB 3; Length 2843;
Best Local Similarity 20.6%; Pred. No. 0.0017;
Matches 124; Conservative 66; Mismatches 235; Indels 176; Gaps 25;

QY 30 TSFSDLLASPLDNNKP--PGQ-----GLSERTGSGVPEKFTPTPPSLPLSPPIPSPPSS 80
DB 2207 SEISGQMKQLOANMPSISGRTHIHPGVNRSSSTSPVSKKGPKLTKTPASKSPSEGT 2266

QY 81 YFAIPPGGLSP---AELLDSPVLLNSSNI-----LPSPTTGAFVAQSF 119
DB 2267 ATTSPRGAKSVKSEL--SPVARQTQIGSSSKAPSRSGRSDTSPSPAQPLSRPIQSP 2324

QY 120 NWKSSSGGNQOIVKEEDKSFNFSPQTRSGPPASSTATYQSSNVTVQTQ-OPMSFQEA TK 178
DB 2325 GRNSISPGRNGISPPNKL S---QLPRTSSPSTASTKSGSGKMSYTPSGRQMSQNNLT K 2380

QY 179 QDNFSSGKGMKTKENSSSMQSFPEIASVQTNHNGFQSDYGNYPQSQTLRSRDDGYN 238
DB 2381 QTGLS-----KNASSIP--RSESASKGLNQNN-----GNANKKVELSRMSS----- 2421

QY 239 WRKYGQKQVKGSENPRSYKCTY-----PNCPTKKKVERS LDGQITIVYKGTNNHPKPN 294

QY 2422 -TKSGSESDBSERPVLVROSTFIKEAPSLRKLKLEAS-----FESLSPSSRPAS 2473

QY 295 TRNNSNSSLAIPHNSIRTEIPDQSYATHGSGQMS-AATPNSISISIGDDDDFEQSSQ 353
DB 2474 PTRSQATPVLS-----PSLDDMSLSTHSSVQAGWRKLPNLSPTI----- 2515

QY 354 KCXSGGDEYDEPDAR-----WKIEGENEGMAGPSRTVREP 392
DB 2516 -----EYNDGRP-AKRHDIAHSHSPRLPINRSGTWKREHSHSSLPVST----- 2563

QY 393 RVVVQTTSDIDLDGVRWKYQKVVKGPNPRSYKCTHPGCPVRKHVERASHDLRAV 452
DB 2564 -----WRTGSSSILSASSESEKAKSED---EKHV-----NSI 2595

QY 453 ITTYEGKHNDVPAARG-----SGHSVNRPMNPNNASNTNTAATSVRLLPVIHQ 502
DB 2596 SGTQSKENQ--VSAKGTWRKIKENEFSPNTSTQTVSSGATNGAESKTLIYQAPAVSK 2653

QY 503 S-----DMSLQNRQSQAPPEQCSPTLEMLQSPGSGFSGFNGPMQSVNQQQLSDN 554
DB 2654 TEDVWVRIEDCPINNPRSGRSPTGNTPPVIDSV-----SEKANPNIKDSKDNQAKQN 2705

QY 555 V 555
DB 2706 V 2706

RESULT 14

US-08-450-582-7
; Sequence 7, Application US/08450582
; Patent No. 6114124
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY

;; APPLICANT: GRODEN, JOANNA
;; APPLICANT: HEDGE, PHILIP J.
;; APPLICANT: JOSLYN, GEOFF
;; APPLICANT: KINZLER, KENNETH
;; APPLICANT: MARKHAM, ALEXANDER F.
;; APPLICANT: NAKAMURA, YUSUKE
;; APPLICANT: THLIVERIS, ANDREW
;; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
;; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
;; NUMBER OF SEQUENCES: 102
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Banner & Witcoff, Ltd.
;; STREET: 1001 G Street, NW
;; CITY: Washington
;; STATE: D.C.
;; COUNTRY: USA
;; ZIP: 20001-4598
;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patent in Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/450,582
;; FILING DATE:
;; CLASSIFICATION: 435
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/452,655
;; FILING DATE: 23-MAY-1995
;; APPLICATION NUMBER: US 08/289,548
;; FILING DATE: 12-AUG-1994
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 07/741,940
;; FILING DATE: 08-AUG-1991
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Kagan, Sarah A.
;; REGISTRATION NUMBER: 32,141
;; REFERENCE/DOCKET NUMBER: 1107.49964
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: 202-508-9100
;; TELEFAX: 202-508-9299
;; INFORMATION FOR SEQ ID NO: 7:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 2843 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: single
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; HYPOTHETICAL: YES
;; ANTI-SENSE: NO
;; US-08-450-582-7

Query Match 5.0%; Score 154; DB 3; Length 2843;
Best Local Similarity 20.6%; Pred. No. 0.0017;
Matches 124; Conservative 66; Mismatches 235; Indels 176; Gaps 25;

QY 30 TSFSDLLASPLDNNKP--PGQ-----GLSERTGSGVPEKFTPTPPSLPLSPPIPSPPSS 80
DB 2207 SEISGQMKQLOANMPSISGRTHIHPGVNRSSSTSPVSKKGPKLTKTPASKSPSEGT 2266

QY 81 YFAIPPGGLSP---AELLDSPVLLNSSNI-----LPSPTTGAFVAQSF 119
DB 2267 ATTSPRGAKSVKSEL--SPVARQTQIGSSSKAPSRSGRSDTSPSPAQPLSRPIQSP 2324

QY 120 NWKSSSGGNQOIVKEEDKSFNFSPQTRSGPPASSTATYQSSNVTVQTQ-OPMSFQEA TK 178
DB 2325 GRNSISPGRNGISPPNKL S---QLPRTSSPSTASTKSGSGKMSYTPSGRQMSQNNLT K 2380

QY 179 QDNFSSGKGMKTKENSSSMQSFPEIASVQTNHNGFQSDYGNYPQSQTLRSRDDGYN 238
DB 2381 QTGLS-----KNASSIP--RSESASKGLNQNN-----GNANKKVELSRMSS----- 2421

QY 239 WRKYGQKQVKGSENPRSYKCTY-----PNCPTKKKVERS LDGQITIVYKGTNNHPKPN 294

Db 2422 -TKSSGESDRSRPVLVRQSTFIKEAPSTLRKLEESAS-----FESLSPSRPAS 2473
QY 295 TRSNSNSSSLAIPHSNIRTEIPDOSYATHSGQOMDS-AATPENSISIGDDDDFQSSQ 353
Db 2474 PTRSQATPVLS-----PSLPDMSLSTHSSVQAGGWRKLPNLSPTI-----2515
QY 354 KCKSGGDEYDEBPAAR-----WKIEGENEGMSAPGSRVREP 392
Db 2516 -----EYNDGRP-AKRHDIAHSHSESPSRLPINRSGTWKREHSHSSSLPRVST----2563
QY 393 RVVVQTTSDIDLDDGVRWKYQKVKVGNPNRPSYKCTHPCGCPVKHVERASHDLRAV 452
Db 2564 -----WRTGSSSSILSASSESEKAKSED---EKHV-----NSI 2595
QY 453 ITTYEGKHNHDVPAARG-----SGSHSVNRPMPNNAHNTTAATSVRLLPVIHQ 502
Db 2596 SGTQSKENQ--VSAKGTWRKIKENEFSPNTNSTQTVSSGATNGAESKTLIIYQMAFAVK 2653
QY 503 S-----DNSLQNRQSQAPPEGQSPFTLEMLQSPGSGFGGNPMOSYVNOQQQLSDN 554
Db 2654 TEDVMVRIEDCPINNPRSGRSPGTNTPPVIDSV-----SEKANPNIKSKDNQAKQN 2705
QY 555 V 555
Db 2706 V 2706

RESULT 15

US-08-449-731-2
; Sequence 2, Application US/08449731
; Patent No. 6413727
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; ANAND, RAKESH
; CARLSON, MARY
; GRODEN, JOANNA
; HEDGE, PHILIP J.
; JOSLYN, GEOFF
; KINZLER, KENNETH
; MARKHAM, ALEXANDER F.
; NAKAMURA, YUSUKE
; THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Allegretti, LTD
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/449,731
; FILING DATE: 25-May-1995
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/289,548
; FILING DATE: 12-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.46943
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 2843 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-08-449-731-2
Query Match 5.0%; Score 154; DB 4; Length 2843;
Best Local Similarity 20.6%; Pred. No. 0.0017;
Matches 124; Conservative 66; Mismatches 235; Indels 176; Gaps 25;
QY 30 TFSDLLASPLDNKLP--PQG-----GLSRTSGVGVKPKFSTPPSPPLPLSPSPSS 80
Db 2207 SEISGQMKQPLQANPISIRGRMTIHIPGVNRSSSTSPVSKGPKLTPASKPSSEGQT 2266
QY 81 YFAIPPLGLSP---AELLDSPVLLNSNI-----LPSPTTCFAVQSF 119
Db 2267 ATTSPRGAKPSVKSEL--SPVARQTSIGGSKAPSRSGRSDTSPRPAQQLSRFIQSP 2324
QY 120 NMKSSSGNQIIVKEEDKSFNFSFOTRSGPPASSTATYQSSNVTVQTQ--QPMWSFOEATK 178
Db 2325 GRNLSIPGRNGISPPNKL-----QLPRTSSPSTASTKSSGSGKMSYTSFGROMSQQLLYK 2380
QY 179 QDNFSSGKGMKMTENSSSMQSFPEIASVQTNHNGFQSDYGNYPQSQTLRRRDDGYN 238
Db 2381 QTGLS-----KNASSIP--RSESASKGLNQMN-----GNANKKVELSRMSS---2421
QY 239 WRKYGQKQVKGSENPRSYKCTY-----PNCPTKKKVERSLEDGOITEIVYKGTNHPKQN 294
Db 2422 -TKSSGESDRSRPVLVRQSTFIKEAPSTLRKLEESAS-----FESLSPSRPAS 2473
QY 295 TRSNSNSSSLAIPHSNIRTEIPDOSYATHSGQOMDS-AATPENSISIGDDDDFQSSQ 353
Db 2474 PTRSQATPVLS-----PSLPDMSLSTHSSVQAGGWRKLPNLSPTI-----2515
QY 354 KCKSGGDEYDEBPAAR-----WKIEGENEGMSAPGSRVREP 392
Db 2516 -----EYNDGRP-AKRHDIAHSHSESPSRLPINRSGTWKREHSHSSSLPRVST----2563
QY 393 RVVVQTTSDIDLDDGVRWKYQKVKVGNPNRPSYKCTHPCGCPVKHVERASHDLRAV 452
Db 2564 -----WRTGSSSSILSASSESEKAKSED---EKHV-----NSI 2595
QY 453 ITTYEGKHNHDVPAARG-----SGSHSVNRPMPNNAHNTTAATSVRLLPVIHQ 502
Db 2596 SGTQSKENQ--VSAKGTWRKIKENEFSPNTNSTQTVSSGATNGAESKTLIIYQMAFAVK 2653
QY 503 S-----DNSLQNRQSQAPPEGQSPFTLEMLQSPGSGFGGNPMOSYVNOQQQLSDN 554
Db 2654 TEDVMVRIEDCPINNPRSGRSPGTNTPPVIDSV-----SEKANPNIKSKDNQAKQN 2705
QY 555 V 555
Db 2706 V 2706

Search completed: January 20, 2004, 16:18:25
Job time : 25 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 20, 2004, 16:19:31 ; Search time 38 Seconds
(without alignments)
3094.123 Million cell updates/sec

Title: US-09-890-811B-10
Perfect score: 3050
Sequence: 1 MASSSGSLDTSASANSFTNF.....FSSRTKEPRDDMFLESLIC 575

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 762491 seqs, 204481190 residues

Total number of hits satisfying chosen parameters: 762491

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1605	52.6	560	9	US-09-810-264-10
2	1145	37.5	519	16	US-10-278-536-182
3	906	29.7	393	15	US-10-278-173-88
4	899	29.5	557	11	US-09-934-455-326
5	899	29.5	557	12	US-10-302-267-198
6	888.5	29.1	278	9	US-09-810-264-28
7	884	29.0	521	15	US-10-278-173-36
8	884	29.0	521	16	US-10-278-536-34
9	861.5	28.2	514	12	US-10-225-068-140
10	861.5	28.2	514	15	US-10-295-403-82
11	858	28.1	577	9	US-09-810-264-24
12	852	27.9	687	11	US-09-934-455-18
13	852	27.9	687	15	US-10-295-403-76
14	848.5	27.8	513	15	US-10-295-403-78
15	753	24.7	266	12	US-10-259-165-64

16	742	24.3	568	11	US-09-934-455-10
17	742	24.3	568	12	US-10-225-068-70
18	643.5	21.1	623	12	US-10-259-165-196
19	625.5	20.5	406	9	US-09-810-264-12
20	619.5	20.3	413	9	US-09-810-264-14
21	618.5	20.3	583	9	US-09-810-264-38
22	606	19.9	430	9	US-09-810-264-16
23	460.5	15.1	258	9	US-09-810-264-32
24	353	11.6	173	9	US-09-810-264-22
25	318	10.4	191	11	US-09-934-455-54
26	316.5	10.4	528	11	US-09-934-455-406
27	315.5	10.3	318	11	US-09-934-455-362
28	315.5	10.3	318	15	US-10-278-173-42
29	310.5	10.2	553	11	US-09-533-029-90
30	310.5	10.2	553	11	US-09-934-455-360
31	309.5	10.1	326	11	US-09-934-455-62
32	309.5	10.1	536	11	US-09-934-455-58
33	309.5	10.1	536	15	US-10-278-173-40
34	309.5	10.1	536	16	US-10-278-536-78
35	305.5	10.0	528	12	US-10-225-068-126
36	300	9.8	337	11	US-09-934-455-296
37	300	9.8	337	15	US-10-278-173-90
38	295	9.7	102	9	US-09-810-264-18
39	288	9.4	277	12	US-10-225-068-144
40	280.5	9.2	147	11	US-09-533-029-16
41	280.5	9.2	147	15	US-10-278-173-64
42	272.5	8.9	317	11	US-09-934-455-12
43	268	8.8	195	11	US-09-934-455-460
44	264.5	8.7	489	11	US-09-533-029-60
45	259.5	8.5	179	11	US-09-934-455-56

ALIGNMENTS

RESULT 1

US-09-810-264-10
; Sequence 10, Application US/09810264
; Patent No. US200207675A1
; GENERAL INFORMATION:
; APPLICANT: Crane, Virginia C.
; APPLICANT: Famodu, Omolayo O.
; APPLICANT: Hu, Xu
; APPLICANT: Lu, Guihua
; APPLICANT: Zhang, Lingyu
; TITLE OF INVENTION: WRKY Transcription Factors and Methods
; TITLE OF INVENTION: of Use
; FILE REFERENCE: 1183
; CURRENT APPLICATION NUMBER: US/09/810,264
; CURRENT FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: US 60/190,467
; PRIOR FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10
; LENGTH: 560
; TYPE: PRT
; ORGANISM: Helianthus annuus
; US-09-810-264-10

Query Match 52.6%; Score 1605; DB 9; Length 560;
Best Local Similarity 57.0%; Pred. No. 1.4e-106;
Matches 335; Conservative 77; Mismatches 120; Indels 56; Gaps 16;

Qy 13 SANSFTNFTFTSTHPTTTSFSDLLASPLDNKKPQ-----GGLSERTGSGVPKFKSTPPP 67
Db 2 SFSSSGITLETTPSPSPFSFMSFSFSDPPPTTGLAARIAERVGSGIPKFKSIPPP 61
Qy 68 SLPLSPPTSPSSVFAIPGLSPAEILLDSPLVILNSNILLPSPTTGAFVAQSEFNWKSSSG- 126
Db 62 SLPIPSPAVSPSSVFAIPAGLSPAEILLDSPLVILNSNILLPSPTTGAFVFNWKNLNGN 121
Qy 127 --GNQIVKEEDKSFNSFQTRSGPPASSTATYQSSNVTVTQQPWSFOEATKQDNFSS 184

Db 122 FHNEHSIKKEQSLADFSRPLQHHPT-----QIWNQ---KQIQDQ 164
Qy 185 GKGMKTSNSSSQSPSPETASVQTHNSNGFQSDYGNYPPOS-----QTLRRSDGYN 238
Db 165 EKSLTQSGHPPMQSPSPETATQDSNQSQSQFQSDYTNSSSNFNQTLQKSEGYN 224
Qy 239 WRKYGQKVGKSNPRSYKCTPNCPTKKKVERSLDGOITEIVYKGTNNHPKPNTRN 298
Db 225 WRKYGQKVGKSNPRSYKCTPNCPTKKKVERSLDGOITEIVYKGTNNHPKPNTRN 294
Qy 299 SSNSSS-----LAIPHNSIRTEIPDOSYATHGSGQMSDAAATPENSISIGDDFEQSSQK 354
Db 285 SSSASNTLQMSQASSNHDVDPQSYVSHGSGQVDSVTPPENSISVSGDDEFDS--- 341
Qy 355 CKSGGD--EYDEDEPAKRWKIEGENEGMS-APGSRVREPRVVQTTSDIDILDDGYRW 411
Db 342 -RSGDGVTVDEDEPAKRWKV-SENEGIMIGTKTVREPRVVQTTSDIDILDDGYRW 399
Qy 412 RYKGVKVGKSNPRSYKCTPNCPTKKKVERSLDGOITEIVYKGTNNHPKPNTRN 471
Db 400 RYKGVKVGKSNPRSYKCTPNCPTKKKVERSLDGOITEIVYKGTNNHPKPNTRN 459
Qy 472 SHSVNRPMPNNSNHTNTAATSVRLLPVH-QSDNSL--QNORSQAPP--EGQSPFTLEM 526
Db 460 HRLQASTLSNNAWSMT-----TKPMALSHYQVDSNVDPTRGPRYPSPSENQAPFTLEM 513
Qy 527 LQSPGSGFGGFGNPMQSYVNOQOLSDNVFSSRTKEEPRDDMFLESLL 574
Db 514 LQSSDNPKYSRFENALKSNYNEHN-SRRTST-TKEEPRDDMFLESLL 559

RESULT 2

US-10-278-536-182
; Sequence 182, Application US/10278536
; Publication No. US20030131386A1
; GENERAL INFORMATION:
; APPLICANT: Samaha, Raymond
; APPLICANT: Heard, Jacqueline
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Pineda, Omaira
; APPLICANT: Reuber, Lynne
; APPLICANT: Riechmann, Jose-Luis
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Keddle, James
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Pilgrim, Marsha
; APPLICANT: Adam, Luc
; APPLICANT: Broun, Pierre
; TITLE OF INVENTION: STRESS-INDUCED POLYNUCLEOTIDES
; FILE REFERENCE: MEI-011
; CURRENT APPLICATION NUMBER: US/10/278,536
; CURRENT FILING DATE: 2002-10-22
; PRIOR APPLICATION NUMBER: 60/125,814
; PRIOR FILING DATE: 1999-03-23
; NUMBER OF SEQ ID NOS: 238
; SOFTWARE: Patentin ver. 2.1
; SEQ ID NO 182
; LENGTH: 519
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: G915
US-10-278-536-182
Query Match 37.5%; Score 1145; DB 16; Length 519;
Best Local Similarity 43.6%; Pred. No. 1,1e-73;
Matches 260; Conservative 82; Mismatches 154; Indels 100; Gaps 15;
Qy 1 MASSGSLD--TSASANSFTFTFTTHPFTTFSDDLASPLDNNKPPQGGLSERTGSG 57
Db 1 MAASFLTMDNSRTRQNMNGSANSSQSGRTSTSSLEDL-----E 39

Qy 58 VPKFKSTPPPSLPLSPPIPSPPSSYFAIPPLGSLPAELLDSPVLLNSS-NILPSPPTTGAFVA 116
Db 40 IKFERSFASSISISPSLVSPTCF-----SPSLFLDSPAFVSSSANVLASPTTGALI- 92
Qy 117 QSNFKWSSGGGQOIVKEEDKFSN-----FSQTRSGPPASSSTATYOSSNVTVQTP 170
Db 93 -----TWVNTQKGINEGDKSNNNFNLFDFSPHTQSSGVGAPTTTTTTTTTTNS- 144
Qy 171 WSPQBATKDNFSSGKGMKMTSSSQSFSPETASVQTHNSNGFQSDYGNYPPOSQTL 230
Db 145 -SIFOSQEQOKKNGEQMSQTETRNQAVS-----YNGRE-----Q 180
Qy 231 RRSDDGYNRWKYQKQVKGSENPSPSYKCTPNCPTKKKVERSLDGOITEIVYKGTNNHP 290
Db 181 RKGEDGYNRWKYQKQVKGSENPSPSYKCTPNCPTKKKVERSLDGOITEIVYKGTNNHP 240
Qy 291 KPQNTNRNNSSSSL-----AIPHNSIRTEIPDOSYATHGSGQMSDAAATPENSISI 343
Db 241 KPQSTRSSSSSTFSAVYNASLDHNRQASSDQPNNSNPFQSDSGFGMQEEDNTTSDV 300
Qy 344 GDDDEQSSQCKSGGDEYDEDEPAKRWKIEGENEGMSAPGSRVREPRVVQTTSDID 403
Db 301 GDDFEQGS-SIVSRDEEDCGSEPEAKRWKGNETNGNGGSGKTVREPRVVQTTSDID 359
Qy 404 ILDDGYNRWKYQKQVKGSENPSPSYKCTPNCPTKKKVERSLDGOITEIVYKGTNNHP 463
Db 360 ILDDGYNRWKYQKQVKGSENPSPSYKCTTIGCPVKHVERASHDMRAVITTEGKHND 419
Qy 464 VPAARGSGSHSVNRPMPNNSNHTNTAATSVRLLPVHOSDNLQNRQAPPEGOSPT 523
Db 420 VPAARGSG-YATNRAPODSSSVPIRPAATA-----GHSNYTTSSQAPYT 462
Qy 524 LEMIQ-----SPGSGFGGFGNPMQSYVNOQOLSDNVFSSRTKEEPRDD-MELESLL 574
Db 463 LQMLHNNNTNTGPGFYAMNNNNNNNLQTCQNFVGGGFSRAKEEPEETSFDFSPM 518

RESULT 3

US-10-278-173-88
; Sequence 88, Application US/10278173
; Publication No. US20030061637A1
; GENERAL INFORMATION:
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Broun, Pierre
; APPLICANT: Riechmann, Jose-Luis
; APPLICANT: Pineda, Omaira
; APPLICANT: Zhang, James
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Pilgrim, Marsha
; APPLICANT: Keddle, James
; APPLICANT: Heard, Jacqueline
; APPLICANT: Reuber, Lynne
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Adam, Luc
; APPLICANT: Samaha, Raymond
; TITLE OF INVENTION: POLYNUCLEOTIDES FOR ROOT TRAIT ALTERATION
; FILE REFERENCE: MEI-009
; CURRENT APPLICATION NUMBER: US/10/278,173
; CURRENT FILING DATE: 2002-10-21
; PRIOR APPLICATION NUMBER: US/09/533,392
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 60/125,814
; PRIOR FILING DATE: 1999-03-23
; NUMBER OF SEQ ID NOS: 177
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 88
; LENGTH: 393
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: G177
US-10-278-173-88

PRIOR APPLICATION NUMBER: 60/124,278
PRIOR FILING DATE: 1999-03-11
PRIOR APPLICATION NUMBER: 60/129,450
PRIOR FILING DATE: 1999-04-15
PRIOR APPLICATION NUMBER: 60/135,134
PRIOR FILING DATE: 1999-05-20
PRIOR APPLICATION NUMBER: 60/144,153
PRIOR FILING DATE: 1999-07-15
PRIOR APPLICATION NUMBER: 60/161,143
PRIOR FILING DATE: 1999-10-22
PRIOR APPLICATION NUMBER: 60/162,656
PRIOR FILING DATE: 1999-11-01
NUMBER OF SEQ ID NOS: 218
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 198
LENGTH: 557
TYPE: PRT
ORGANISM: Arabidopsis thaliana
FEATURE:
OTHER INFORMATION: G1382
US-10-302-267-198

Query Match 29.5%; Score 899; DB 12; Length 557;
Best Local Similarity 40.4%; Pred. No. 5e-56;
Matches 226; Conservative 64; Mismatches 165; Indels 104; Gaps 20;

QY 55 GSGVPKFKSTPPSLPLSPSPSSYFAIPPGLSPAELLDSPVLLNSSNLLPSPTGA-113
DB 32 GGGGARYKLMSPAKL-----PIRSTDTITPPGLSPTSFLESFVFINIKPEPSPTGSL 86
QY 114 FVAQSPNWKSSG-----GNOQIVKEEDKSFNSFOTRSGPPASSTATQSSNVTVQTQ 168
DB 87 FKRPVHISASSSSYGRGHQNTFTQKS-SEFEFR-----PASNMVYAELEK--IRSE 139
QY 169 QPMSFQ-----EATQDNFSGKG-----MMKTENSSSMQSFSPSEIASVQTNHS 212
DB 140 PPVHFQCGHSGHSPSSISDAAGSSSELGRPTPPCQMTPTSSDIPAGSDQESIQTSQN 199
QY 213 NGFQSDYGNVPPQSOTLSRSDGYNWRKYQKQVKGSENPNSRYKCTYPCNCTKKKVER 272
DB 200 DS-----RGSTPSILADDGYNWRKYQKQVKGSENPNSRYKCTYPCNCTKKKFER 249
QY 273 SLGQITEIVYKTHHPKQNTNRNNSSS-----LAIPHNSIRTEIPDQSVYATHGSG 328
DB 250 SHGQITDIYKTHDPKQPGRNNSGCGAAQEERLDKYPSTGRDEKSGVY--NLSN 307
QY 329 QMDSAATPENSISIGDDDFQSSQCKSGGDEYDEPDPAKRWKIEGENEGMSAPGSR 388
DB 308 PNEQTGNPEVPPISASDDGEEAAS--NRNKDEPDDDDFSKRRMEGAME--ITPLVKP 363
QY 389 VREPRVVQTSDDILDDGYRWKRYQKQVKGSENPNSRYKCTYPCNCTKKKVERASHD 448
DB 364 IREPRVVVQTLSEVDILDDGYRWKRYQKQVKGSENPNSRYKCTYPCNCTKKKVERASHD 423
QY 449 LRAVITYEGKHNDHVAAGSGSHSVN--RP-----MPNNSNHTN 488
DB 424 PKAVITYEGKHNDHVTSSKSNHETQPRFDEDTISLNLGVGSISSDGPVHASNE-- 481
QY 489 TAATSVALLPVHQSDNSLQNSQAPPEGSPPTLEMLQSPGSGFGNPMQSVYVQ 548
DB 482 -----HQHQN--QQLVQNTHPNGV-----NFRFV-HASPMSSYYAS 514
QY 549 QQLSDNVFSSR-TKEPRD 566
DB 515 LNSGMNQYQORETKNETQN 533

RESULT 6

US-09-810-264-28
Sequence 28, Application US/09810264
Patent No. US2002007675A1
GENERAL INFORMATION:
APPLICANT: Crane, Virginia C.

APPLICANT: Famodu, Omolayo O.
APPLICANT: Hu, Xu
APPLICANT: Lu, Guibua
APPLICANT: Zhang, Lingyu
TITLE OF INVENTION: WRKY Transcription Factors and Methods
FILE REFERENCE: 1183
CURRENT APPLICATION NUMBER: US/09/810,264
PRIOR FILING DATE: 2001-03-16
PRIOR APPLICATION NUMBER: US 60/190,467
PRIOR FILING DATE: 2000-03-17
NUMBER OF SEQ ID NOS: 43
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 28
LENGTH: 278
TYPE: PRT
ORGANISM: Triticum aestivum
US-09-810-264-28

Query Match 29.1%; Score 888.5; DB 9; Length 278;
Best Local Similarity 52.4%; Pred. No. 1.1e-55;
Matches 184; Conservative 41; Mismatches 47; Indels 79; Gaps 11;

QY 230 SRRSDGYNWRKYQKQVKGSENPNSRYKCTYPCNCTKKKVERSL-DGQITEIVYKTHN 288
DB 2 NKQVEDGYNWRKYQKQVKGSENPNSRYKCTYPCNCTKKKVERSLADGRITQIVYKGAHD 61
QY 289 HPKQNTNRNNSSSS-LAIPHNSIRTEIPDQSVYATHGSDSAATPENSISIGDDDD 347
DB 62 HPKPPSTRNNSGCAAVIAEDHTN-----GSEHSGTPENSISVTFGDDDE 105
QY 348 FEQSSQCKSGGDEYDEPDPAKRWKIEGENEGMS-APGS--RTVREPRVVVQTSDDIDI 404
DB 106 ADNGA-----EPETKGRKEHGDNEGSSGTCACVKPVRPRVLVWQTSDDIDI 152
QY 405 LDDGYRWKRYQKQVKGSENPNSRYKCTYPCNCTKKKVERASHDLRAVITYEGKHNDH 464
DB 153 LDDGFRWRKYQKQVKGSENPNSRYKCTYPCNCTKKKVERASHDLRAVITYEGKHNDH 212
QY 465 PAARGSGSHSVNPMNNSNHTNTAATSVRLLPVHQSDNSLQNSQAPPEGSPPTL 524
DB 213 PVGRGAGA--SRALPTSSSD-----SSVVVVCFA-----AAGQAPYTL 248
QY 525 EMLQSPGSGFGSGFGNPMQSVYVQQLSDNVFSSRTKEPRDDMFLESLLC 575
DB 249 EMLANPAA-GRGYA-----AKDEPRDDMFVESLLC 278

RESULT 7

US-10-278-173-36
Sequence 36, Application US/10278173
Publication No. US20030061637A1
GENERAL INFORMATION:
APPLICANT: Jiang, Cai-Zhong
APPLICANT: Broun, Pierre
APPLICANT: Riechmann, Jose-Luis
APPLICANT: Pineda, Omaira
APPLICANT: Zhang, James
APPLICANT: Yu, Guo-Liang
APPLICANT: Pilgrim, Marsha
APPLICANT: Keddle, James
APPLICANT: Heard, Jacqueline
APPLICANT: Reuber, Lynne
APPLICANT: Ratcliffe, Oliver
APPLICANT: Adam, Luc
APPLICANT: Samaha, Raymond
TITLE OF INVENTION: POLYNUCLEOTIDES FOR ROOT TRAIT ALTERATION
FILE REFERENCE: MBI-009
CURRENT APPLICATION NUMBER: US/10/278,173
CURRENT FILING DATE: 2002-10-21
PRIOR APPLICATION NUMBER: US/09/533,392
PRIOR FILING DATE: 2000-03-22
PRIOR APPLICATION NUMBER: 60/125,814


```
; NAME/KEY: DOMAIN
; LOCATION: (227)...(285)
; OTHER INFORMATION: Conserved domain
; FEATURE:
; NAME/KEY: DOMAIN
; LOCATION: (407)...(465)
; OTHER INFORMATION: Conserved domain
US-10-225-068-140

Query Match      28.2%; Score 861.5; DB 12; Length 514;
Best Local Similarity 38.3%; Pred. No. 2.2e-53;
Matches 217; Conservative 62; Mismatches 139; Indels 149; Gaps 21;

QY 31 SFDLLASPLDNKPPQGLS-----ERTGSGV-----PKFKSTPPSLPLSPPP 75
DB 60 SFDLLAGWSSPATAAAAAATASDYQRLGEGTSSSGDVPFRKQNRPTGLMISQSQ 119
QY 76 ISPSYFAIPGGLSPALLDSPVLLNSSNLPSTTGAFVAQSFNWKSSSGNQIIVKEE 135
DB 120 -SP-SMFTVPGLSPALLDSPFLG---LFSVQGSY-----GMTHQQAQA- 162
QY 136 DKSFNFSFOTRSGPPASATYQSSNVTVQTPQPSFQATKQDNFSSGKGMKMTENS 195
DB 163 -----VTAQAVOANANMQPTEYP-----PPSQVQSFSSGQAQIPTSA 202
QY 196 SMQSFSPETASVQTNHNSGFQSDYGNYPPOSQTLRSRSDGYNWRKYQKQVKGSE 255
DB 203 PAQRETSDVITIE-----HRSQPLNVDPKADDDGYNWRKYQKQVKGSEFPRS 250
QY 256 YKCTYPCNCTKKKVERSLOGQTEIYVKGTHNHPQNTRR--NSSNSSSLAIPHNSIR 314
DB 251 YKCTNPGCPVKKVERSLOGQTEIYKQHNHPPQNTKRGKNDTANI---NGSSIN 307
QY 315 TEIPDOSYATHSGQMSAATPENSISIGDDDFEQSSQK-----KSGDEY----- 362
DB 308 N-----NRGSELGASQFQTNSSNKTREQHEAVSQATTTTEHLSEASDGSEVNGET 359
QY 363 -----DEDEPAKRWKIEGE-NEGMSAPGSRVTREPRVVVQTTSDIDLDGGRWKYQ 416
DB 360 DVREKDENEPDKRRSTEVRISEPAASHRTVTEPRIIVQTTSEVDLDDGGRWKYQ 419
QY 417 KVKGNPNRPSYKCTHPGCPVKRKHVERASHDLRAVITYEGKHNHDPVPAARGSGSHVN 476
DB 420 KVKGNPNRPSYKCTHPGCGVRKHVERAATDPKAVVTYEGKHNHDLPAK--SSH--- 475
QY 477 RMPNNASNHTNTAATSVRLLPVTHQSDNSLQNSORQAPEGOSPTFLEMLQSPGSGFS 536
DB 476 -----AAAAQLRP-----DN-----RPG----- 489
QY 537 GFGNPMQSYVNOOQLSDNVFSSRTKEE 563
DB 490 GLAN-----LNQOQOQOQFVARLRKEE 511

RESULT 11
US-09-810-264-24
; Sequence 24, Application US/09810264
; Patent No. US20020076775A1
; GENERAL INFORMATION:
; APPLICANT: Crane, Virginia C.
```

```
; APPLICANT: Famodu, Omolayo O.
; APPLICANT: Hu, Xu
; APPLICANT: Lu, Guihua
; APPLICANT: Zhang, Lingyu
; TITLE OF INVENTION: WRKY Transcription Factors and Methods
; TITLE OF INVENTION: of Use
; FILE REFERENCE: 1183
; CURRENT APPLICATION NUMBER: US/09/810,264
; CURRENT FILING DATE: 2001-03-16
; PRIOR APPLICATION NUMBER: US 60/190,467
; PRIOR FILING DATE: 2000-03-17
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 24
; LENGTH: 577
; TYPE: PRT
; ORGANISM: Glycine max
US-09-810-264--24

Query Match      28.1%; Score 858; DB 9; Length 577;
Best Local Similarity 35.8%; Pred. No. 4.5e-53;
Matches 216; Conservative 78; Mismatches 175; Indels 134; Gaps 20;

;
QY 41 DNKKPQGGLSERTGSGVPKFKSTPPPSLPSPSPSSYFAIPGCLSPAELLDSPVLL 100
Db 10 DPNRPSAADAAPAGA-RYKLLSPAKLPISRSPC-----VTISPCLSPTSFLESPVLL 63
QY 101 NSSNIIPLSPPTG--AFVAQSFNWKSSSGGNQOI-----VKEEDKSFNFSFOIR 147
Db 64 SNNKVEPSPPTGSLSLHLOTAYGSMTSAASATFPVLCASIAIPLMRNLAF--FEFKPH 121
QY 148 SGPPASSTATYQSSNVTVQTPQWMSFOEATKQNFSSGK-----GMKNTSSSMQ-- 198
Db 122 SG-----SNMVPADFNDHASEKSTQID--SQKAQAPDSSALVKNESASPSNEL 168
QY 199 SFSPEIASVQTNHNSGQSDYGNYPQSQ-----TLRSRSDGQYXW 239
Db 169 SLSPVKWIAQEAASARVEGDLDELNPRSNITTTGLQASQVDRNGSLTVAARVSDDGYNW 228
QY 240 RYKGQKQVKGSENPSPRYKCTYPCPTKKKVERSLOGQITVYKGTNNHHPKQNTNRNS 299
Db 229 RYKGQKHVKGSEFPSPRYKCTHPNCEVKKLFERSHDGQITVYKGTNNHHPKQNTNRNS 288
QY 300 SNSSSLAIPHSNIRTEIPQ-----SYATHGGQMDSAATPENSISIGDDDFEQSSQK 354
Db 289 A-CTIMSVOEDRDKASLTSRDCKGSMNCGGSHLAEPCGPPELLPATNDGDLGLVL 347
QY 355 CKSGGDEYDEPDAPKRWKIEGENGMS--APGSRVTPRVPVVVQTTSDIDLDGQYWR 412
Db 348 SNRNDEVDVDDPFSEKSKMD--VGIADITPVVVPKPIREPVVVVQTLSEVDILDDGQYWR 404
QY 413 KYGQKQVKGSENPSPRYKCTHPGCPVKHVERASHDLRAVITYYEGKNNHDPVPAARGSGS 472
Db 405 KYGQKQVKGSENPSPRYKCTHPGCPVKHVERASHDPAVITYYEGKNNHDPVPAARNS-C 463
QY 473 HSNVRMPNNAASHT-----NTAATSVRL-----PVTHOS 503
Db 464 HDWAGFA--SASGOTRVPRESDTISLDLGMGISPAEANTNSQGRMNLSEFGDSQIHTS 521
QY 504 DNSLQNRQAPPEGQSPFTLEMLQSPGFGFGNPMQSYVYVQOQSDNVSFSTRKEE 563
Db 522 NSNFK-----FVHTTAPGYFGV-----LNNNSNPYGSKEN 552
QY 564 PRD 566
Db 553 PSD 555

RESULT 12
US-09-934-455-18
; Sequence 18, Application US/09934455
; Publication No. US20030121070A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Adam, Luc
; APPLICANT: Creelman, Robert
; APPLICANT: Dubell, Arnold
; APPLICANT: Heard, Jacqueline
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Keddle, James
; APPLICANT: Pilgrim, Marsha
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Reuber, Lynne
; APPLICANT: Riechmann, Jose Luis
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Pineda, Omaira
; TITLE OF INVENTION: Genes for Modifying Plant Traits IV
; FILE REFERENCE: MBI-0025
; CURRENT APPLICATION NUMBER: US/09/934,455
; CURRENT FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/227439
; PRIOR FILING DATE: 2000-08-22
; PRIOR APPLICATION NUMBER: MBI-0022
; PRIOR FILING DATE: 2001-11-16
; PRIOR APPLICATION NUMBER: MBI-0023
; PRIOR FILING DATE: 2001-04-17
; NUMBER OF SEQ ID NOS: 516
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 18
; LENGTH: 687
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-934-455-18

Query Match      27.9%; Score 852; DB 11; Length 687;
Best Local Similarity 37.1%; Pred. No. 1.5e-52;
Matches 229; Conservative 63; Mismatches 146; Indels 180; Gaps 24;

;
QY 47 QGGLSE-----RTGSGVPKFKSTPPPSLPSPSPSSYFAI-----PGLSP 90
Db 72 RGLSRIARAARAFNAPR-----LNTNIRNTDFSIDNSLRSPCLTITSSPCLSP 121
QY 91 AELLDSPVLLNSNIIPLSPPTGAFVAQSFNWKSSSGGN--OOVKEE--DKSFSNFSQ 145
Db 122 ATLLSPVFLNSPLAQSPPTG-----KFFFLPGVNGNALSKAKDEFFDDIGASFSEH 176
QY 146 TRSGPPASSTATYQSSNVTVQTPQWMSFOEATKQDNFSSGKMMKTENSSSMQSFPEIA 205
Db 177 ----PVSRSSSF-----FQGTTEMSVDYGNYNRSSHQSAEVEKPGSE 218
QY 206 SVQTNHNSGQSDYGNYPQSQTL-----SRSD-----DGY 237
Db 219 NIESNLYGIETD--NQNGNKTSDDVTNTSLTVDHQEEEEQRGDSNAGAPAEADGY 276
QY 238 NMRKYGQKQVKGSENPSPRYKCTYPCPTKKKVERSLOGQITVYKGTNNHHPKQNTNR 297
Db 277 NMRKYGQKQVKGSENPSPRYKCTNPNCOVKKVEREGHITVYKGTNNHHPKQNTNR 336
QY 298 N-----SSNSSLAIPHSNIRTEIPQSYAT 324
Db 337 SGQVQDGTGEVQEQOQORDSAAATVSCNNTQQCGGSNNV---ESGSTRFEYGNQSGSI 393
QY 325 HG--SQMDNS-----AATPENSISIGDDDFEQSSQKKS-----GGDEYDEDEDAKR 371
Db 394 QAQTGQYSGSDGPPVVVVDASSTFSDNDEDDDRGTHGSVSLGVDGGGGGGGDESESKR 453
QY 372 WKIEGENGMSAPGSTVREPVVVQTTSDIDLDGQYWRKYGQKQVKGSENPSPRYK 431
Db 454 RKLFAFAAEMSG--STRAIREPVVVQTTSDVDILDDGQYWRKYGQKQVKGSENPSPRYK 512
QY 432 THPGCPVKHVERASHDLRAVITYYEGKNNHDPVPAAR-----GSGSHSVNRMPNNAASHT 487
Db 513 TAPGCTVRKHVERASHDLKSVITYYEGKNNHDPVPAARNSHGGGDSGN-----GNSG 565
QY 488 NTAATSVRLPVTHQSDNSLQNRQAPPEG-----QSPFTLEMLQSPGSP----- 533
Db 566 GSAAVS-----HHVHNGHSEPPRGRPDROVTTNNQSPFS-----RPFSPQPHL 609
```


Db 95 SVGGGSGGVDPRFRQSRPTGLMITQPP-----GMFTVPPGLSPATLLDSPFFG---- 146
Qy 106 LPSPPTTCFAVAQSNFNWSSGGNQIIVKBEKDKSFSNFSFOTRSGPPASSTA-TYQSSNVT 164
Db 147 LFPPLQTF-----GMTHQALAQVTAQAVQGNVH 177
Qy 165 VQTQPMFSQBATKQDNFNSGKGMKMTENSSSMQSFSPETASVQ-TNHSNGFQSDYG--N 221
Db 178 MQSQSQSEYSSSTQOQQOQQOQASLTEIPSFSSAPRSQIRASVQETSQOQRETSEISVPE 237
Qy 222 YPQSQTLSRRSDGYRWRYGQKVKGSENPRSYKCTYPNCPCTKKKVERSLEDGQITEI 281
Db 238 HRSQPQADKPADDGYRWRYGQKVKGSDPPRPSYKCTHPACPVKKKVERSLEDGQVTEI 297
Qy 282 VYKGNHHPQPNTRRNSNSSSLAIPHNSIRTEIPDQSYATHGSGQMDSAATPENSSI 341
Db 298 IYKQGNHHELPQKGNNGSCKS-----SDIANQ-FQTSNSLNKSKEDQETSQV 346
Qy 342 SIGDDDFEQSSQK----CKSGGDEYDEPDAKRWKIEGENEGMSAPGSRVTRVPRVVQ 397
Db 347 TTTEQMSSEASDSEVGVNAETSVGERHEDEPDPKRNTVVRVSEPVASSHRTVTEPRIIVQ 406
Qy 398 TTSDIDLDDGYRWRYGQKVKGNNPNRPSYKCTHGGCPVYKVERASHDLRAVITYE 457
Db 407 TTSEVDLLDDGYRWRYGQKVKGNNPYRPSYKCTTPDCGVKXKVERAATDPKAVVTYE 466
Qy 458 GKNNHDPVPAARGSGSHSVNRPMPNPNASNTHTNT 489
Db 467 GKNNHDPVPAARTS-SHOLR---PNNQHN-TST 493

RESULT 15
US-10-259-165-64
; Sequence 64, Application US/10259165
; Publication No. US20030135888A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; APPLICANT: Chang, Hur-song
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiyaki
; APPLICANT: Kreps, Joel
; APPLICANT: Moughamer, Todd
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; TITLE OF INVENTION: GENES THAT ARE MODULATED BY POSTTRANSCRIPTIONAL GENE SILENCING
; FILE REFERENCE: 70030-NP
; CURRENT APPLICATION NUMBER: US/10/259,165
; PRIOR FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: US 60/368,327
; PRIOR FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; NUMBER OF SEQ ID NOS: 782
; SOFTWARE: PatentList.pl version 3.0.4 (C) 2001 Syngenta
; SEQ ID NO 64
; LENGTH: 266
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-259-165-64

Query Match 24.7%; Score 753; DB 12; Length 266;
Best Local Similarity 51.3%; Pred. No. 5.2e-46;
Matches 162; Conservative 28; Mismatches 88; Indels 58; Gaps 8;
Qy 267 KKKVERSL-DQQITEIYVKGTHNHPKPNTRRNSNSSSLAIPHNSIRTEIPDQSYATH 325
||||||| |||:||||| ||||| :|||:| :|

Db 2 KKKVERSLADGRITQIYVKGANNHPKPLSTRNASSCATAA-----ACADDLAAPG 52
Qy 326 GSGQMDSAATPENSSISIGDDDFEQSSQKCKSGGDEYDEPDAKRWKIEGENEGMS--- 382
Db 53 AGADQYSAATPENSSVTFGDEADNASHSEG-----DEPEAKEWKEDADNEGSGGM 105
Qy 383 --APGSRVREPRVVQTTSDIILDDGYRWRYGQKVKGNNPNRPSYKCTHPPGCPVRK 440
Db 106 GGGAGGKPVREPRLVVQTLSDIILDDGFRWRKYGQVYKGNPNRPSYKCTTVGCPVRK 165
Qy 441 HVERASHDLRAVITTYEGKHHNDVPAARGSGSHSVNRPMPNPNASNTHTNTAATSVRLPVI 500
Db 166 HVERASHDTRAVITTYEGKHHNDVPGVGGGGGRAPAP-----PTSGAIRP--- 213
Qy 501 HQSDNSLQNRQSQAPPEGQSPPTLEMLQSP-GSFGSGFGNPMQSYVNOQQQLSDNYFSSR 559
Db 214 -----SAAAAQQQPYTLEMLPNPAGLYGGYGAG-----AGGAAFP 250
Qy 560 TKEEPRDDMPLESLLC 575
Db 251 TKDERRDDLFVESLLC 266

Search completed: January 20, 2004, 16:25:08
Job time : 40 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 20, 2004, 16:16:54 ; Search time 20 Seconds
(without alignments)
2764.850 Million cell updates/sec

Title: US-09-890-811B-10
Perfect score: 3050
Sequence: 1 MASSSGSLDTSASANSFTNF.....FSSRTKEPRDDMFLESLLC 575

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB ID	Description
1	1822	59.7	549	2 S51529	SPF1 protein - swe
2	1331	43.6	514	2 S72443	DNA-binding protei
3	1139.5	37.4	512	2 T02498	probable WRKY-type
4	1032.5	33.9	402	2 S61413	DNA-binding protei
5	906	29.7	393	2 T00575	probable WRKY-type
6	888	29.1	571	2 T08930	hypothetical prote
7	884	29.0	309	2 T48481	SPF1-like protein
8	848.5	27.8	513	2 C84447	probable WRKY DNA-
9	840	27.5	528	2 T52092	DNA-binding protei
10	794.5	26.0	487	2 G86272	hypothetical prote
11	778.5	25.5	509	2 JC6203	sp8 binding protei
12	742	24.3	568	2 T05060	hypothetical prote
13	565.5	18.5	296	2 S72444	DNA-binding protei
14	561.5	18.4	349	2 E84790	probable WRKY-type
15	556.5	18.2	487	2 F84462	transcription fact
16	520	17.0	1895	2 T06609	disease resistance
17	488.5	16.0	782	2 B85362	hypothetical prote
18	316.5	10.4	528	2 C85056	probable DNA-bindi
19	306	10.0	548	2 T49948	hypothetical prote
20	301	9.9	327	2 T04919	DNA-binding protei
21	300	9.8	337	2 A84913	probable WRKY-type
22	299.5	9.8	458	2 T49114	hypothetical prote
23	295.5	9.7	252	2 B86422	FIN18.10 protein -
24	292.5	9.6	304	2 T09357	hypothetical prote
25	291.5	9.6	287	2 B96717	unknown protein, 3
26	288	9.4	294	2 T48026	hypothetical prote
27	288	9.4	471	2 E86322	hypothetical prote
28	275	9.0	374	2 G96704	probable DNA bindi
29	272.5	8.9	317	2 T00500	probable elicitor

30	264.5	8.7	489	2 B85022	probable DNA-bindi
31	259	8.5	349	2 S61414	DNA-binding protei
32	258.5	8.5	403	2 T02003	probable DNA-bindi
33	257.5	8.4	410	2 C84638	probable WRKY-type
34	257	8.4	341	2 S72445	DNA-binding protei
35	257	8.4	353	2 T09887	DNA-binding protei
36	251.5	8.2	324	2 T10685	DNA-binding protei
37	245.5	8.0	380	2 C84710	hypothetical prote
38	235.5	7.7	302	2 A96841	hypothetical prote
39	235.5	7.7	387	2 C96720	hypothetical prote
40	235	7.7	97	2 A84899	probable WRKY-type
41	228.5	7.5	624	2 T05090	hypothetical prote
42	227	7.4	304	2 T05587	hypothetical prote
43	220.5	7.2	197	2 E84606	probable WRKY-type
44	220.5	7.2	421	2 H86431	T518.10 protein -
45	220.5	7.2	427	2 T00465	probable WRKY-type

ALIGNMENTS

RESULT 1

S51529
SPF1 protein - sweet potato
C;Species: Ipomoea batatas (sweet potato)
C;Date: 15-Jul-1995 #sequence_revision 01-Sep-1995 #text_change 21-Jul-2000
C;Accession: S51529
R;Ishiguro, S.; Nakamura, K.
Mol. Gen. Genet. 244, 563-571, 1994
A;Title: Characterization of a cDNA encoding a novel DNA-binding protein, SPF1, that re-

O.
A;Reference number: S51529; MUID:95058910; PMID:7969025

A;Accession: S51529

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-549 <158>

A;Cross-references: EMBL:D30038; NID:g484260; PIDN:BAA06278.1; PID:g484261

Query Match 59.7%; Score 1822; DB 2; Length 549;

Best Local Similarity 62.0%; Pred. No. 7e-101; Indels 68; Gaps 14;
Matches 369; Conservative 66; Mismatches 92;

QY	1	MASSGSLDTSAN	SFNFTSTHPTFTSFDLLASPLDNNKPPQGGI	-----SRTGS	56
Db	1	MAASGOTIDAPTASSS	---FSTASSFMSFTDLLASDAYSGSVSRGLGDR	IAERTGS	57
QY	57	GVPKFKSTPPSLPLSPPI	SPSSYPFAIPGLSPAELDSPVLLNSNLPSTTGA	FVA	116
Db	58	GVPKFKSLPPSLPLSSPAVSPSSYPFAIPGLSPELLDSPVLLSSNLPSTTGT	FPA	117	
QY	117	QSFNWKSSGGNQIIVKEEDKSFNSFOTRSGPPASSTATYQSSNVTVTQCPWFS	FOEA	176	
Db	118	QTFNWKNDNSAOEDVKEEBKGYDPDFSQTNS	---ASMTLNVEDSK	---	160
QY	177	TKQDNFSSGKGMKMTENSSMQSFPEIASVQTNHNSN	--GFQSDYGNV--PPOSQTL	--SR	231
Db	161	-----RKDELNSLQSLPPVTSTQWSSQNGSGSYSEYNNQCCPP	QOTLREQR	207	
QY	232	RSDGYNWRKYQKQVKGS	ENPRSYKCTYPCPKKKVRS	LDGOITEIVYKGTNNHPK	291
Db	208	RSDGYNWRKYQKQVKGS	ENPRSYKCTHPNCPTKKKVERALDGOITEIVYKGAHNHPK	267	
QY	292	PONTRENS	---SNSSSLAI PHSNSIRTEIPQOSVATHSGQMD	SAATPENSISIGDD	347
Db	268	POSTRRSSSTASSASTLAAQSYNAPASDVPDQSYWNSNGQMD	SVATPENSISVGDD	327	
QY	348	FQSSQCKSGGDEYDE	DEPDARWKIEGENGMSAPGSRTRVREPRVVQTTSD	IIDLD	407
Db	328	FQSSQCKSGGDEYDE	DEPDARWKIEGENGMSAPGSRTRVREPRVVQTTSD	IIDLD	387
QY	408	GYRWRYGQVKVGNPN	PRSYKCTHPGCPVRKHVERASHDLRAVITTYEGKHNH	VPAA	467
Db	388	GYRWRYGQVKVGNPN	PRSYKCTSGQCPVRKHVERASHDLRAVITTYEGKHNH	VPAA	447

Db 484 TQCNFVGGSFRAKEBNEETSFDDSPM 511

RESULT 4

S61413

DNA-binding protein ABF1 - wild oat (fragment)

C:Species: Avena fatua (wild oat)

C>Date: 27-Apr-1996 #sequence_revision 13-Mar-1997 #text_change 21-Jul-2000

C:Accession: S61413

R:Rushon, P.J.; Macdonald, H.; Huttly, A.K.; Lazarus, C.M.; Hooley, R.

Plant Mol. Biol. 29, 691-702, 1995

A:Title: Members of a new family of DNA-binding proteins bind to a conserved cis-element

A:Reference number: S61413; MUID:96128012; PMID:8541496

A:Accession: S61413

A:Status: preliminary

A:Molecule type: mRNA

A:Residues: 1-402 <RUS>

A:Cross-references: EMBL:Z48429; NID:g1159876; PID:CAA88326.1; PID:g1159877

Query Match 33.9%; Score 1032.5; DB 2; Length 402;

Best Local Similarity 53.2%; Pred. No. 3.6e-54;

Matches 225; Conservative 48; Mismatches 105; Indels 45; Gaps 15;

QY 168 QPWSFOEATKQNFSSGKGMKMTENSSMQSPSEIASVOTNHSNGFQSDYGNYPQSQ 227

Db 2 QQPWGYQQQ-----PGAM--DAGANAASF--APAVQATSSB--MAPSGGVYRQTH 46

QY 228 TLSRRSDGYNWRKYQKQVKGSENPRSYKCTPNCPTKKKVERSILDGOITEIVYKGT 287

Db 47 SQRSSDDGYNWRKYQKQVKGSENPRSYKCTFPNCPKKKVTESIEGQITEIVYKGT 106

QY 288 NHPKQNTNRSSNSSSLAIPHNSIRTEIPDQSYATHGSGQMSA--ATPENSISIGD 345

Db 107 NHAQLSTRSGGGGGGAA-----QVLSGGGGDASEHSGFAMSGAPVSTPENSASFSD 162

QY 346 DDFEQSSQCKS--GGDYDEDEPDAKRWKIEGENEGM--SAPGSRTRVPRVVVQTTSDID 403

Db 163 DETGASSPRAGNVGDDLDDEPDKKWDGDEGSGNSMAGNKTREPRVVVQTTMSDID 222

QY 404 ILDDGYRWKYQKQVKGSENPRSYKCTHPGCPVRKHVERASHDLRAVITTYEGKHND 463

Db 223 ILDDGYRWKYQKQVKGSENPRSYKCTTVGCPVRKHVERASHDLRAVITTYEGKHND 282

QY 464 VPAARGSGSHSVNRPMPNNSN---HTNTAATSVRLLP---VIHQSNLSLQNSQAPPEG 518

Db 283 VPAARSA--ALYRPAARAAADSMSTSQYTANQOQPSMTYTQTSNAACTQYAPRPG 340

QY 519 -----QSPFTLEMLQSPSGFSGS--GFGNPMQSYVN-----QQQLSDNVFSSRTKEEPRDDM 568

Db 341 FGSONQSGFGFN-----GSFGSAGFDNPTGYSYMSHQOQOQOQNDAMQASGAKEEPREDM 395

QY 569 FLE 571

Db 396 FFQ 398

RESULT 5

T00575

probable WRKY-type DNA binding protein [imported] - Arabidopsis thaliana

N:Alternate names: probable DNA-binding protein T27E13.1

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 23-Mar-2001

C:Accession: T00575; B4706

R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul

submitted to the EMBL Data Library, May 1998

A:Description: Arabidopsis thaliana chromosome II BAC T27E13 genomic sequence.

A:Reference number: Z14178

A:Accession: T00575

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-393 <ROU>

A:Cross-references: EMBL:AC004165; NID:g3150396; PID:g3150397

A:Experimental source: cultivar Columbia

R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L. euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487; PMID:10617197

A:Accession: B84706

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-393 <STO>

A:Cross-references: GB:AE002093; NID:g2347191; PIDN:AAC16930.1; GSPDB:GN00139

C:Genetics:

A:Gene: At2g30250; T27E13.1

A:Map position: 2

A:Introns: 79/3; 117/3; 294/2; 348/2

Query Match 29.7%; Score 906; DB 2; Length 393;

Best Local Similarity 45.4%; Pred. No. 1.1e-46;

Matches 209; Conservative 59; Mismatches 98; Indels 94; Gaps 13;

QY 29 TTTSFDLL-ASPLDNNKPPQ-----GGLSERTSGVPKPKSTPPSLPLSPPTIS 77

Db 3 STSFTDLLGSSGVCDEDEDLRVSGSPGGYYPERTGSLPKFKAQPP-----PLPIS 57

QY 78 PSYFAIPPLGLSPAEILLDSFVLLNSNILPSPPTGAFVAQSFNWKSSSGGNQOIVKEEDK 137

Db 58 QSSH-----NFTFSDYLDSPLLSSSLISPTTGTTFPLQGFNGTTN----- 99

QY 138 SFSNFSFQTRSGPPASSTATYSSNVTVTQOPWSPQEQATKQDNFSSGKMKMTENSSM 197

Db 100 NHDSPWQIQSQP-----SNASSALOETYGVDHEK----- 131

QY 198 QSFSP-EIASVOTNHSNGFQSDYGNYPPOSQTLRSRSDGYNWRKYQKQVKGSENPRSY 256

Db 132 QEMIPNEIATQNNQSGFGERQI-KLP--AYWVRNSNDYGWRKYGQKVKSENPRSY 188

QY 257 YKCTYPCPTKKKVERSLDQITEIVYKGTNHPKQNTNRSSNSSSLAIPHS----- 310

Db 189 FKCTYPCVSKKIVETASDQITEIYKGTNHPKPEFTKRPSSQSS---LPSSVNGRRL 244

QY 311 ---NSIRTIPOQSYATHGSGQMSAATPENSISIGDDDFEQSSQCKSGGDEYDEDEP 367

Db 245 FNPASVYSPHDSQ-----ENSSISFDYSDLEQKSPKSYGEIDEDEEQP 289

QY 368 DAKRWKIEGENSGMAGPAGSRTRVPRVVVQTTSDIDLDGYNWRKYQKQVKGSENPRSY 427

Db 290 EMKRMKREGDEGMSIEVSGKVPKRVVVTISDIDLDGYNWRKYQKQVKGSENPRSY 349

QY 428 YKCTHPGCPVRKHVERASHDLRAVITTYEGKHNDVPA 467

Db 350 YKCTFGGCVKQKQVRSADRAVITTYEGRHNDIPTA 389

RESULT 6

T08930

hypothetical protein T15N24.90 - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C>Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 22-Oct-1999

C:Accession: T08930

R;Bevan, M.; Zimmermann, W.; Grueneisen, A.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; Ma;

submitted to the Protein Sequence Database, May 1999

A:Reference number: Z16518

A:Accession: T08930

A:Molecule type: DNA

A:Residues: 1-571 <BEV>

A:Cross-references: EMBL:AL078465; GSPDB:GN00062; ATSP:T15N24.90

A:Experimental source: cultivar Columbia; BAC clone T15N24

C:Genetics:

A:Gene: ATSP:T15N24.90

A:Map position: 4

A:Introns: 76/3; 90/3; 142/3; 311/1; 363/2; 415/2

Query Match 29.1%; Score 888; DB 2; Length 571;

Best Local Similarity 39.8%; Pred. No. 2.1e-45;
Matches 229; Conservative 62; Mismatches 162; Indels 122; Gaps 22;

QY 55 GSGVFKSTPPSLPLSPPISSYFAIPGLSPAELDSPVLLNSNIL----- 106
DB 32 GGGARYKLMSPAKL-----PISRSTDTITPPGLSPTSFLESFVI--SNIKTVFFKTS 84

QY 107 -----PSPTTCA-FVAOSFNWSSG-----GNOQIVKEEDKGSFNSFQTRSGPPA 152
DB 85 CLLIDQPEPFTTGLSPKRPVHISASSSYTGGRFHQNTFTQKS-SEFEFR-----PPA 139

QY 153 SSTATYQSSNVTVTQOPWFSQ-----EATKQDNFSSGKG-----MMKTENSSS 196
DB 140 SNMVAELGK--IRSEPPVHFQCGCHSHSPSSIDAAGSSSELSPPTPCQMTPTSSD 197

QY 197 MQSFSPFIASVQTNHNSGFOSDYGNYPPQSQTLRRSDGYNWRKYQKQKGVKGNENPRSY 256
DB 198 IPAGSDQEEISQTSNDSS-----RGSTPSILADDGYNWRKYQKQKGVKGNENPRSY 247

QY 257 YCTYPCNPTKKKVERSILGQITEIVYKTHHPKQPNTRNSNSS-----LAIPHNS 312
DB 248 YKTHPCNCEYKGFERSHDGQITDIYKGDHPKQPGRRNSGMAAQERLQKYPST 307

QY 313 IRTEIPQSVATHSGQMDSAATPENSSISIGDDDFQSSQCKSGDGEYDEDEPDARW 372
DB 308 GDEKSGSVY--NLSNPNQETGNPEVPPISASDDGGEAAS--NRNKDEPDDDPFSKR 363

QY 373 KIEGENGMSAPGSRVREPRVVQTTSDIDLDGYNWRKYQKQKGVKGNENPRSYKCT 432
DB 364 RMGAMB--ITPLVKPIREPRVVQTLSEVDILDDGYNWRKYQKQKGVKGNENPRSYKCT 421

QY 433 HPGCPVKHVERASHDLRAVITTYEGKHNDVPAARGSHSVN--RP----- 478
DB 422 AHGCPVKHVERASHDPAVITTYEGKHNDVPAARGSHSVN--RP----- 478

QY 479 -----MPNASHNHTNTAATSVRLLPVHQSDNSLQNRQAPGEGOSPTLEMLQSPGS 532
DB 482 VGTSSDQPNHASN-----HQHQN--QQLVQNTHPNGV-----N 513

QY 533 FGSGFGNPMQSYVNOQLSDNVFSSR-TKEEPRD 566
DB 514 FRFV-HASPMSSYVSLNSGWNQYQRETKNETQN 547

RESULT 7
T48481
SPFI-like protein - Arabidopsis thaliana
N:Alternate names: protein T28J14.40
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C:Accession: T48481
R:Bavan, M.; Murphy, G.; Ridley, P.; Hudson, S.; Bancroft, I.; Mewes, H.W.; Rudd, S.; Le
submitted to the Protein Sequence Database, April 2000
A:Reference number: 224493
A:Accession: T48481
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-309 <BEV>
A:Cross-references: EMBL:AL163652
A:Experimental source: cultivar Columbia; BAC clone T28J14
C:Genetics:
A:Map position: 5
A:Introns: 51/3; 93/3; 211/2; 254/2
A>Note: T28J14.40

Query Match 29.0%; Score 884; DB 2; Length 309;
Best Local Similarity 47.7%; Pred. No. 1.7e-45;
Matches 198; Conservative 30; Mismatches 55; Indels 132; Gaps 13;

QY 58 VPKFKSTPPSLPLSPPISSYFAIPGLSPAELDSPVLLNSNILSPPTTGAFVAQ 117
DB 10 VPKFKTATP-----SPLPLSPSPYFTMPPLGTLTADFLDPLLTSSNILSPPTTGTPPAQ 64

QY 118 SFNWKSSSG-----NQOIVKEEDKGSFNSFQTRSGPPASSTATYQSSNVTVTQOPWFSQ 174
DB 65 SLNY--NNGLLIDKNEIKVEDTT-----PP-----LFLPSMVT-----QPLQL 102

QY 175 EATKQDNFSSGKMGKMTENSSSQSFSPFIASVQTNHNSGFOSDYGNYPPQSQTLRRSD 234
DB 103 DLFGSEIMSNK-----TSD 117

QY 235 DGVNWRKYQKQKGVKGNENPRSYKCTYPCNPTKKKVERSIL-DGQITEIVYKTHHPKQ 293
DB 118 DGVNWRKYQKQKGVKGNENPRSYKCTYPCNCLTKKKVETSLVKGQMIETVYKSHNHPKQ 177

QY 294 NTRRNSNSSSLAIPHNSIRTEIPQSVATHSGQMDSAATPENSSISIGDDDFQSSQ 353
DB 178 STKSSSTAIA--AHQNS-----SNGD----- 197

QY 354 KCKSGGDEYDEPDARWKIENEGMSAPGSRVREPRVVQTTSDIDLDGYNWRK 413
DB 198 -----GKDICEDETEAKRWKE-EN-----VKEPRVVQTTSDIDLDGYNWRK 241

QY 414 YGQKVVKGPNPRSYKCTHPCGPKVKHVERASHDLRAVITTYEGKHNDVPAAR 468
DB 242 YGQKVVKGPNPRSYKCTFTGCFVKHVERAFQDPKSVITTYEGKHQIPTR 296

RESULT 8
C84447
probable WRKY DNA-binding protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: C84447
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
euse, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: C84447
A:Molecule type: DNA
A>Status: preliminary
A:Residues: 1-513 <STO>
A:Cross-references: GB:AE002093; NID:g4335764; PIDN:ADI7441.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2G03340
A:Map position: 2

Query Match 27.8%; Score 848.5; DB 2; Length 513;
Best Local Similarity 41.0%; Pred. No. 4.1e-43;
Matches 210; Conservative 52; Mismatches 169; Indels 81; Gaps 15;

QY 5 SGLSDTSASANSFTNFTFTTHPMTTSFSDLLASPL-----DNNKPPQG 48
DB 36 SGGVGSPGPMTLVSNLFS--DPDEFKSFQOLLGAMASPAASAAVAVATAHQTPTS 94

QY 49 GLSERTGSG---VPKFKSTPPSLPLSPPISSYFAIPGLSPAELDSPVLLNSNI 105
DB 95 SVGDGGSGGDVDPFRKQSRPTGLMITQPP---GMFTVPGLSPATLLDSPSFFG--- 146

QY 106 LPSPTTGAFVAOSFNWSSGNGNQIIVKEEDKGSFNSFQTRSGPPASSTA-TYQSSNV 164
DB 147 LPSPLQGT-----GMTHQOALLAQVTAQAVQGNVH 177

QY 165 VQTPQPSQEATKQDNFSSGKMGKMTENSSSQSFSPFIASVQ-TNHSNGFOSDYG--N 221
DB 178 MQOSQOSEYPSSTQOQQOQQOQASLTETIFSFSAQPSQIRASVQETSGQRETSISVPE 237

QY 222 YPQSQSTLRRSDGYNWRKYQKQKGVKGNENPRSYKCTYPCNPTKKKVERSILGQITEI 281
DB 238 HRSQPQNAKPADDDGYNWRKYQKQKGVKGNENPRSYKCTYPCNPTKKKVERSILGQVTEI 297

QY 282 VYKGNHHPKQPNTRNSNSSSLAIPHNSIRTEIPQSVATHSGQMDSAATPENSSI 341
DB 298 IYKQHNHELPOKRGNNNGSCKS-----SDIANQ--FQTSNSLNSKSKRDQETSQV 346

C;Species: Cucumis sativus (cucumber)
C;Date: 11-Apr-1997 #sequence_revision 09-May-1997 #text_change 21-Jul-2000
C;Accession: JC6203
R;Kim, D.J.; Smith, S.M.; Leaver, C.J.
Gene 185, 265-269, 1997
A;Title: A cDNA encoding a putative SPFL-type DNA-binding protein from cucumber.
A;Reference number: JC6203; MUID:97208883; PMID:9055825
A;Accession: JC6203
A;Molecule type: mRNA
A;Residues: 1-509 <KIM>
A;Cross-references: GB:141134; NID:927024; PIDN:AAC37515.1; PID:927025
C;Comment: This protein is a DNA-binding protein consisting of a monomer. It is involved
C;Genetics:
A;Gene: s271
C;Keywords: DNA binding

Query Match 25.5%; Score 778.5; DB 2; Length 509;
Best Local Similarity 39.9%; Pred. No. 5.9e-39;
Matches 196; Conservative 63; Mismatches 167; Indels 65; Gaps 15;

QY 31 SFDLLASPLDNNKPPGGISERTGSGVPKFKSTPPSLPL-----SPP---PISPSSY 81
DB 61 SFSQLLAGAMAS---PMAMGFFGTGS-TPNYAKDGPASELEFGMKQSKPVLNLVARSPL 116
QY 82 FAIPPGSLPAELLDSPVLLNSSNLTPTTGAFVAQSFNWKSSGGNQQIVKEEDKSFN 141
DB 117 FSVPPGSLPGLLNSP-----GFYPPQSPFGMSHQQAALAQTAQAALANSH 162
QY 142 FSPQTRSGPPASSTATYQSSNVTVTQOPWSFOEATKQDNFSSGKGMMKTENSSMQSFS 201
DB 163 MHMQ-----QAEYOHSSVPAPT-EPLVRDPSFSLDDASQALAIPTSDTKLIAES 212
QY 202 PETASVQTNUN-GFQSDYGNYPQSQTLRSRSDGYNWKYQKQKQVKGSENPRSYKCT 260
DB 213 TEV-----SHSDRYQP-----PPPHGSKDPAADDGYNWKYQKQKVGSEFPRSYKCT 262
QY 261 YPNCPTKKKVERSLDGQITETVYKGTNHPKPQNTRENSNSSSLAIPHNSNIRTEIPDQ 320
DB 263 HLNCPRKKKEGLPDGTEITLIYKQNNHPPPPANKARDNI EPAGCTNS-LIKPEGCLQ 321
QY 321 SYATHGSGQMSAATPENSSISIGDDDFEQQSKCKSGGDEYDEDEPAKRWKIEGENEG 380
DB 322 NQA--GILNKSSENVQLGSSDSEGRADTEIT-----DDRDEDEPNKQINDAGTSG 371
QY 381 MSAPGSTVREPRVVQVTTSDIDLDGYNWKYQKQKQVKGSENPRSYKCTHPGCPVRK 440
DB 372 V-ALSHKLTLEPKIIVQTRSEVDLDGYNWKYQKQKQVKGSENPRSYKCTSAGCNVRK 430
QY 441 HVERASHDLRAVITTYEGKHNHVPAAARGSGSHSVNRPMPNPNASNHTNTAATSRLPLVI 500
DB 431 HVERSSDTSKAVVTTYEGKHNHVPAAARNSSHHTV-----NNTVHHIKPLKVAQKHPLL 485
QY 501 HQSDNSLQNR 511
DB 486 KEMEFQNDQR 496

RESULT 12
T05060
hypothetical protein M3E9.130 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-Jun-1999
C;Accession: T05060
R;Bevan, M.; Vandenbol, M.; Jallet, C.; Portetelle, D.; Hoheisel, J.; Mewes, H.W.; Mayer
submitted to the Protein Sequence Database, March 1999
A;Reference number: 215396
A;Accession: T05060
A;Molecule type: DNA
A;Residues: 1-568 <BEV>
A;Cross-references: EMBL:AL022223
A;Experimental source: cultivar Columbia; BAC clone M3E9
C;Genetics:
A;Map position: 4

A;Introns: 106/3; 339/2; 392/2
A;Note: M3E9.130

Query Match 24.3%; Score 742; DB 2; Length 568;
Best Local Similarity 37.3%; Pred. No. 9.9e-37;
Matches 193; Conservative 55; Mismatches 119; Indels 150; Gaps 16;

QY 86 PGLSPAELLDSPVLLNSSNLTPTTGAFVAQSFNWKSSGGNQQIVKEEDKSFNFRFQ 145
DB 88 PGLSPATLLESFVFL--SNPLLSPTTKL-----SSVPSDKAKAELFDDITSLAFQ 137
QY 146 TRSGPPASSTATYQSSNVTVTQOPWSFOEATKQDNFSSGKGMMKTENSSMQSFSPEA 205
DB 138 TISGGLDPT-----NIALEPDDSDQYEERQ-----LGGLG-----DSMACCAP--- 176
QY 206 SVQTNHNSGFSQSDYGNYPQSQTLRSRSDGYNWKYQKQKQVKGSENPRSYKCTVPNCP 265
DB 177 -----ADGGYNWKYQKQKQVKGSEYPRSYKCTHPNCE 209
QY 266 TKKKVERSLDGQITETVYKGTNHPKPQNTRENSNSSSLA----- 306
DB 210 AKKKVERSPEGHLEIITGDIHHSKPPNRRSGISGSGTQDMQIDATEYEGFAGTNN 269
QY 307 IPHNSIRTEIPDQSYA-----THSGQMSAATPENSSISIGDDDFEQQSKCKSG 358
DB 270 IEWTSFVSAELEYGSHSGSMQVQNGTHQFGYDAAA---DALYRDENEEDRTSHMSVSL 325
QY 359 GHEYDEDEPAKRWKIEG---ENEGMSAPGSTVREPRVVQVTTSDIDLDGYNWKYQ 415
DB 326 TYDGEVESESKRRLKAEAYATSG-----STRASREPRVVQVTTSDIDLDGYNWKYQ 381
QY 416 QKVKGNNPRSYKCTHPGCPVRKHVERASHDLRAVITTYEGKHNHVPAAAR-----GS 470
DB 382 QKVKGNNPRSYKCTANGCTVTKHVERASDDFKSVLTITYGKHTVVVPAARNSSHVGA 441
QY 471 GS-----HSVNRMPNPNASNHTNTAATSRLPLPVHQSNSLQNRQAAP 516
DB 442 GSSGTLQGLSATQTNHNHVPMPHRSRSEGLATA-----NSSLDFDQSH--- 485
QY 517 EGQSPFTLEMLQSPSGFSGFGNPMQSVVNOQLSD 553
DB 486 -----LRHPTGFSV-----YIGQSELSD 503

RESULT 13
S72444
DNA-binding protein WRKY2 - parsley (fragment)
C;Species: Petroselinum crispum (parsley)
C;Date: 12-Feb-1998 #sequence_revision 13-Mar-1998 #text_change 21-Jul-2000
C;Accession: S72444
R;Rushton, P.J.; Tovar Torres, J.; Parniske, M.; Wernert, P.; Hahlbrock, K.; Somssich,
EMBO J. 15, 5690-5700, 1996
A;Title: Interaction of elicitor-induced DNA-binding proteins with elicitor response el
A;Reference number: S72443; MUID:97051827; PMID:8896462
A;Accession: S72444
A;Molecule type: mRNA
A;Residues: 1-296 <EUS>
A;Cross-references: EMBL:U58540; NID:g1432057; PIDN:AAC49529.1; PID:g1432058
C;Keywords: DNA binding

Query Match 18.5%; Score 565.5; DB 2; Length 296;
Best Local Similarity 43.0%; Pred. No. 1.3e-26;
Matches 132; Conservative 36; Mismatches 68; Indels 71; Gaps 9;

QY 271 ERLDGOITETVYKGTNHPKPQNTRR-----NSSSSSLAIPHNSIRTEIP 318
DB 1 ERALDGOITETVYKGTNHPKPQPSRRRTAGALISIQEEKAVNASL-----TGQG 51
QY 319 DQSVATHGSGQMSAATPENSSISIGDDDFEQQSKCKSGGDEYDEDEPAKRWKIEGEN 378
DB 52 DTTISQTLA--DQNGTPISSPRGVNADVDGASPLNSVTDDIDNDQDFMKRRRTDVG 109
QY 379 EGMSPGSTVREPRVVQVTTSDIDLDGYNWKYQKQKQVKGNNPRSYKCTHPGCPV 438

[illegible]

RESULT 14

RES001 14
 E84790
 probable WRKY-type DNA binding protein [imported] - Arabidopsis thaliana
 C.Species: Arabidopsis thaliana (mouse-ear cress)
 C.Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
 C.Accession: E84790
 R.Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
 euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
 Nature 402, 761-768, 1999
 A.Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A.Reference number: A84420; MUID:20083487; PMID:10617197
 A.Accession: E84790
 A.Status: preliminary
 A.Molecule type: DNA
 A.Residues: 1-349 <STO>
 A.Cross-references: GB:AE002093; NID:G4056481; PIDN:AAC98047.1; GSPDB:GN00139
 C.Genetics:
 A.Gene: At2g37260
 A.Map position: 2

RESULT 15

RESULT 15
F84462
transcription factor ZAP1 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: F84462
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, M.; Xoo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Rhee, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Eaves, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: F84462
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-487 <STO>
A:Cross-references: GB:AE002093; NID:g4585919; PIDN:AAD25579.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g04880
A:Map position: 2

Search completed: January 20, 2004, 16:20:29
Job time : 23 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 20, 2004, 16:16:52 ; Search time 18 Seconds
(without alignments)
1502.242 Million cell updates/sec

Title: US-09-890-811B-10

Perfect score: 3050

Sequence: 1 MASSSGSLDTSASANSFTNF.....FSRTKEPRDDMFLESLLC 575

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1139.5	37.4	512	1 WR33_ARATH	Q889P5 arabidopsis
2	906	29.7	393	1 WR25_ARATH	O22921 arabidopsis
3	899	29.5	557	1 WR20_ARATH	Q93W70 arabidopsis
4	884	29.0	309	1 WR26_ARATH	Q9C5T3 arabidopsis
5	861.5	28.2	514	1 WRK4_ARATH	Q9X190 arabidopsis
6	852	27.9	687	1 WRK2_ARATH	Q9F977 arabidopsis
7	848.5	27.8	513	1 WRK3_ARATH	Q9ZQ70 arabidopsis
8	742	24.3	568	1 WR34_ARATH	O65590 arabidopsis
9	738	24.2	423	1 WR58_ARATH	Q93WU7 arabidopsis
10	566.5	18.6	429	1 WR44_ARATH	Q9ZUU0 arabidopsis
11	556.5	18.2	487	1 WRK1_ARATH	Q9I377 arabidopsis
12	520	17.0	1895	1 WR19_ARATH	Q98Z67 arabidopsis
13	482	15.8	466	1 WR32_ARATH	F59583 arabidopsis
14	386	12.7	506	1 WR10_ARATH	Q91G05 arabidopsis
15	348	11.4	399	1 WR48_ARATH	Q9FGZ4 arabidopsis
16	318	10.4	218	1 WR12_ARATH	Q93WY4 arabidopsis
17	316.5	10.4	528	1 WR42_ARATH	Q9XEC3 arabidopsis
18	315.5	10.3	318	1 WR28_ARATH	Q8VWJ2 arabidopsis
19	310.5	10.2	553	1 WRK6_ARATH	Q9C5I9 arabidopsis
20	309.5	10.1	326	1 WRK8_ARATH	Q9F126 arabidopsis
21	306	10.0	548	1 WR72_ARATH	Q91XG8 arabidopsis
22	305.5	10.0	480	1 WR61_ARATH	Q8VWV6 arabidopsis
23	303.5	10.0	538	1 WR31_ARATH	Q93WR0 arabidopsis
24	301	9.9	282	1 WR71_ARATH	Q93WV4 arabidopsis
25	300	9.8	337	1 WR23_ARATH	O22900 arabidopsis
26	292.5	9.6	304	1 WR13_ARATH	Q9SVB7 arabidopsis
27	291.5	9.6	287	1 WR68_ARATH	Q9C983 arabidopsis
28	288	9.4	277	1 WR57_ARATH	Q93WV6 arabidopsis
29	280.5	9.2	147	1 WR45_ARATH	Q98763 arabidopsis
30	275	9.0	374	1 WRK9_ARATH	Q9C9I0 arabidopsis
31	272.5	8.9	317	1 WR15_ARATH	O22176 arabidopsis
32	268	8.8	195	1 WR56_ARATH	Q8VWQ4 arabidopsis
33	264.5	8.7	145	1 WR75_ARATH	Q9FYA2 arabidopsis

RESULT 1

WR33_ARATH
ID WR33_ARATH STANDARD; PRT; 512 AA.
AC Q889P5;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Probable WRKY transcription factor 33 (WRKY DNA-binding protein 33).
GN WRKY33 OR AT2G38470 OR T19C21.4.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosidb II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia; TISSUE=Flower;
RA Lippok B., Somssich I.E.;
RT "Arabidopsis thaliana transcription factor WRKY33."
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Rinning C.M., Koo H.L.,
RA Moffat L.S., Cronin L.A., Shen M., Pai G., Van Aken S., Unayam L.,
RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
RA Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,
RA Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
RA Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana."
RL Nature 402:761-768(1999)
CC -!- FUNCTION: Transcription factor. Interacts specifically with the W box (5'-(T)TGAC(C/T)-3'), a frequently occurring elicitor-responsive cis-acting element (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- SIMILARITY: Belongs to the WRKY class I family.
CC -!- SIMILARITY: Contains 2 WRKY domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC EMBL: AF509499; AAC34736.1; -;
CC EMBL: AC004683; ABM14994.1; -;
CC FIR; T02498; T02498.
CC InterPro; IPR003657; WRKY.
CC Pfam; PF03106; WRKY; 2.
CC PROSITE; P550811; WRKY; 2.

ALIGNMENTS

34	264.5	8.7	489	1	WR47_ARATH	Q9ZSI7 arabidopsis
35	259.5	8.5	179	1	WR24_ARATH	Q9FFS3 arabidopsis
36	257	8.4	353	1	WRK7_ARATH	Q9STX0 arabidopsis
37	256	8.4	173	1	WR50_ARATH	Q8VWQ5 arabidopsis
38	251.5	8.2	324	1	WR11_ARATH	Q9SV15 arabidopsis
39	247	8.1	321	1	WR17_ARATH	Q9SJA8 arabidopsis
40	245.5	8.0	380	1	WR21_ARATH	O04336 arabidopsis
41	243	8.0	194	1	WR51_ARATH	Q93WU9 arabidopsis
42	240	7.9	330	1	WR74_ARATH	Q93WU6 arabidopsis
43	236.5	7.8	330	1	WR39_ARATH	Q9S077 arabidopsis
44	236	7.7	109	1	WR43_ARATH	Q8GYL1 arabidopsis
45	235.5	7.7	302	1	WR40_ARATH	Q9SAH7 arabidopsis

```
KW Transcription regulation; DNA-binding; Nuclear protein; Repeat.
FT DOMAIN 123 135 THR-RICH.
FT DNA_BIND 171 235 WRKY 1.
FT DNA_BIND 349 414 WRKY 2.
FT DOMAIN 461 481 ASN-RICH.
SQ SEQUENCE 512 AA; 56457 MW; 8F19CBE41BC18662 CRC64;

Query Match 37.4%; Score 1139.5; DB 1; Length 512;
Best Local Similarity 44.7%; Pred. No. 3.7e-56;
Matches 254; Conservative 80; Mismatches 144; Indels 90; Gaps 15;

QY 40 LDNNKPPQ-----GGLSRTGS-----GVKFKSTPPSLPLSPSPSSYPAP 85
DB 1 MDNSTRQNWNGSANWQQSGRTSTSLDLLEIFKFRSFPAPSSISISPLVSPSTCF--- 57
QY 86 PGLSPABELLPSVLLNS-NILPPTTGAFVAQSFNWKSSGGNQIVKGEDEKFSFN--- 141
DB 58 ---SPSLFLDSPAFVSSSANVLASPTTGALI-----TNVTNKGKINEGDKSNNNFN 106
QY 142 ---PSFOTRSGPPASSTATVQSSNVTVTQOPHSGFQKATKODNFSSGKGMKTNSSMQ 198
DB 107 LDFSFHTQSGVSAPTITTTTTTTTTTNS--SIFOSQEQKKNQSQEQSQETREPNQ 164
QY 199 SFSPEIASVQTNHSGQSDYGNYPQSTLSRRSDGYNWRKYQKQVKGSENPRSYK 258
DB 165 AVS-----YNGRE-----QRKGEGYNWRKYQKQVKGSENPRSYK 201
QY 259 CTYPNCTPKKVERSLLDQITEIVYKTHNHPKQNTNRSSNSL-----AIPHSN 311
DB 202 CTYPNCTPKKVERSLLDQITEIVYKTHNHPKQNTNRSSNSL-----AIPHSN 261
QY 312 SIRTEIPDQSVATHGSGQMSAATPENSISIGDDFEQSQKCKSGDEDEDEPAKR 371
DB 262 QASDQPNNSNPFQSDGFGQEQEDNTTSDSGDDEFEQGS-STVSRDEEDCGSEPAKR 320
QY 372 WKIGENEGSAPGSRVREPRVVQTTSDIDLDGVRWKYQKQVKGSENPRSYK 431
DB 321 WKGNETNGNGGSKVREPRVVQTTSDIDLDGVRWKYQKQVKGSENPRSYK 380
QY 432 THPCPVKRVKVERASHDLRAVITYEGKHNHVDPAARGSGSHSVNRPMPNNASHTTAA 491
DB 381 TTIGCPVKRVKVERASHDMRAVITYEGKHNHVDPAARGSG-YATNRAPODSSSVPIRPA 439
QY 492 TSVLLPVIHQSNLQNSQAPGSGPTLEMLQ-----SPSGFSGFGNPMQSYVN 547
DB 440 IA-----GHSNYTSSQAPYTLQMLHNNNTNTGPGYAMNNNNNNNLQ 483
QY 548 QQQLSDNVFSRTKEPRDD-MFLESLL 574
DB 484 TQQNFVGGGFSRAKEEPEETSFDFSPM 511

RESULT 2
WR25 ARATH STANDARD; PRT; 393 AA.
AC O22921; Q94AT4;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Probable WRKY transcription factor 25 (WRKY DNA-binding protein 25).
GN WRKY25 OR AT2G30250 OR T9D9.6.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
OC NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Ulker B., Kuehnir S., Somersich I.E.;
RA "Arabidopsis thaliana transcription factor WRKY25.";
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.

QY 29 TTTSFSDLL-ASPLDNNKPPQ-----GGLSERTSGVGPVKFSTPPSLPLSPSPIS 77
DB 3 STSFDTLLGSSGVDCCYEDDDLRVSSSGFCGYPTERTGSLPKFKAQPP-----PLPIS 57
QY 78 PSSYFAIPPGLSPAELLDSPVLNNSNLPSTTGAFVAQSFNWKSSGGNQIVKGEDEK 137
DB 78 PSSYFAIPPGLSPAELLDSPVLNNSNLPSTTGAFVAQSFNWKSSGGNQIVKGEDEK 137

SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
MEDLINE=20083487; PubMed=10617197;
Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,
Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.L.,
Moffat K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Unayam L.,
Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,
Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
Venter J.C.;
"Sequence and analysis of chromosome 2 of the plant Arabidopsis
thaliana.";
Nature 402:761-768 (1999).
[3]
SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
"RIKEN Arabidopsis full length cDNA clones (RAFLs) sequenced by the
SSP consortium (Salk/Stanford/PGEU).";
Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
[4]
SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
Yamada K., Chan M.M., Chang C.H., Dale J.M., Hsuan V.W., Lee J.M.,
Onodera C.S., Quach H.L., Tang C.C., Toriumi M., Wong C., Wu H.C.,
Yu G., Yuan S., Chen H., Cheuk R., Jones T., Kim C.J., Nguyen M.,
Palm C.J., Shinn P., Southwick A., Tripp M.G., Wu T., Davis R.W.,
Ecker J.R., Theologis A.;
"Arabidopsis open reading frame (ORF) clones.";
Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
RL "FUNCTION: Transcription factor. Interacts specifically with the W
box (5'-(T)TGAC(C/T)-3'), a frequently occurring elicitor-
responsive cis-acting element (5y similarity).
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- SIMILARITY: Belongs to the WRKY class I family.
CC -1- SIMILARITY: Contains 2 WRKY domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
EMBL; AF418309; AAL13040.1; -
DR EMBL; AC002338; AAC16930.1; -
DR EMBL; AC004165; AAM14918.1; -
DR EMBL; AY062720; AAL32798.1; -
DR EMBL; AY114650; AAM47969.1; -
DR EMBL; AY136318; AAM96984.1; -
DR EMBL; AY045813; AAK76487.2; -
DR EMBL; BT002338; AAN86171.1; -
DR PIR; T00575; T00575.
DR InterPro; IPR003657; WRKY.
DR Pfam; PF03106; WRKY; 2.
DR PROSITE; PS50811; WRKY; 2.
KW Transcription regulation; DNA-binding; Nuclear protein; Repeat.
FT DNA_BIND 160 224 WRKY 1.
FT DNA_BIND 322 387 WRKY 2.
SQ SEQUENCE 393 AA; 44133 MW; 9057D65B69E12A0C CRC64;

Query Match 29.7%; Score 906; DB 1; Length 393;
Best Local Similarity 45.4%; Pred. No. 2.1e-43;
Matches 209; Conservative 59; Mismatches 98; Indels 94; Gaps 13;

QY 29 TTTSFSDLL-ASPLDNNKPPQ-----GGLSERTSGVGPVKFSTPPSLPLSPSPIS 77
DB 3 STSFDTLLGSSGVDCCYEDDDLRVSSSGFCGYPTERTGSLPKFKAQPP-----PLPIS 57
QY 78 PSSYFAIPPGLSPAELLDSPVLNNSNLPSTTGAFVAQSFNWKSSGGNQIVKGEDEK 137
DB 78 PSSYFAIPPGLSPAELLDSPVLNNSNLPSTTGAFVAQSFNWKSSGGNQIVKGEDEK 137
```


Db 58 QSSH-----NFTSDYLDSPLLSSSHLSISPITGTFPLQFGNCTN----- 99
Qy 138 SFNFQTSRGPSPASSTATYQSSNVTVQTPQWFSOEATKQDNFSGGKGMKMTSSSM 197
Db 100 NHDSPWQLOQSP-----SNASSALQETVYGVQDHEKK----- 131
Qy 198 QSFSP-EIASVQTNHNSGPOSDYGVNPPQSCQTSRRSDGYNWRKYGQKQVKGSENPVS 256
Db 132 QEMIPNEIATONNQSFGTQR-I-KIP--AYMVSNSNDGYNWRKYGQKQVKGSENPVS 188
Qy 257 YKTYNCPKPKKVERSLDGOITIVYKVTNHPKPKONTRNSNSSLAIPIHS----- 310
Db 189 FKTYPCVSKKIVETASDQITTEIYKGGNHPKPEFTKRPQSS-----LPSVNGRRL 244
Qy 311 ---NSIRTEPDQSYATHGSGQMSAATPNSISIGDDDFEOSQCKSGGDEYDEDEP 367
Db 245 FNPASVSEPHDQS-----ENSSISFDYSLEQKSPKSEGEIDEDEEQP 289
Qy 368 DAKRWKIEGNEGMSAPGSTVREPRVVVQTTSDIDILDGYRWKYGQKQVKGSENPVS 427
Db 290 EMKRMKREGDEGMSIEVSKGKPEPRVVVQTTSDIDILDGYRWKYGQKQVKGSENPVS 349
Qy 428 YKCTHGPCVRKHVERASHDLRAVITTYEGKNHDVPA 467
Db 350 YKCTFGCGVKKQKVERSAADERAVLTYEGRNHNDIPTA 389

RESULT 3
WR20 ARATH STANDARD; Q94AP6; Q9SU40; PRT; 557 AA.
AC Q93WV0; Q8H1E9; Q94AP6; Q9SU40;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Probable WRKY transcription factor 20 (WRKY DNA-binding protein 20).
GN WRKY20 OR A74G2640 OR T15N24.90.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 1).
RC STRAIN=cv. Columbia; Tissue=Flower;
RT Ulker B., Kushnir S., Somasich I.E.,
RA "Arabidopsis thaliana transcription factor WRKY20".
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
MEDLINE=20083488; PubMed=10617198;
RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
RA Pohl T., Dueterhoeft A., Stiekema W., Enian K.-D., Terryn N.,
RA Harris B., Ansoorge W., Brandt P., Grivell L.A., Rieger M.,
RA Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,
RA Kreis M., Delsen M., Putgdomenech P., Watson M., Schmidheini T.,
RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
RA Vos P., Hohseil J., Zimmermann W., Wedler H., Ridley P.,
RA Langham S.-A., McCullagh B., Bilham L., Robben J.,
RA Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,
RA Braeken M., Weijens I., Voet M., Bastiaens I., Aert R., Defoor E.,
RA Weitzenecker T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
RA Holzer E., Brandt A., Peters S., van Staveren M., Dirks W.,
RA Moeljan P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
RA Bernieris S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
RA De Keyser A., Buysshaert C., Gielens J., Villarroel R., De Clercq R.,
RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McLay K., Mayes R.,
RA Pettett A., Rajandream M.A., Lyne M., Bense V., Rechmann S.,
RA Borkova D., Bloeker H., Scharf M., Grimm M., Loeferl T.-H.,
RA Dose S., de Haan M., Maarse A.C., Schaefer M., Loeferl T.-H.,
RA Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzl A.,
RA Neumann S., Argiriou A., Vitale D., Liguori R., Piravandi E.,

RA Massenot O., Quigley F., Clabaud G., Muendlein A., Feiber R.,
RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
RA Chefor F., Cooke R., Berger C., Monfort A., Casacuberta E.,
RA Gibbons T., Weber N., Vandenbol M., Bagues M., Terol J., Torres A.,
RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse C.,
RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bielek C.,
RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
RA Stonking T., Kalicki J., Graves T., Harmon G., Edwards J.,
RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
RA Nelson J., Spith J., Ryan E., Andrews S., Geisel C., Layman D.,
RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Jordan C.,
RA Ma P., Zhong J., Preston R., Vill D., Shekher M., Matero A., Shah R.,
RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
RA Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
RA Chen E., Maria M., Martienssen R., McCombie W.R.;
RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
RT thaliana".
RL Nature 402:769-777(1999).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC STRAIN=cv. Columbia;
RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RT "RIKEN Arabidopsis full length cDNA clones (RAFLs) sequenced by the
RT SSP consortium (Salk/Stanford/PGEC)".
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC STRAIN=cv. Columbia;
RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Hsuan V.W., Lee J.M.,
RA Quach H.L., Tang C.C., Toriumi M., Wallender E.K., Wong C., Wu H.C.,
RA Yu G., Yuan S., Chen H., Cheuk R., Jones T., Kim C.J., Nguyen M.,
RA Palm C.J., Shinn P., Southwick A., Tripp M.G., Wu T., Davis R.W.,
RA Ecker J.R., Theologis A.;
RT "Arabidopsis open reading frame (ORF) clones".
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Transcription factor. Interacts specifically with the W
CC box (5'-(T)TGAC(C/T)-3'), a frequently occurring elicitor-
CC responsive cis-acting element (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q93WV0-1; Sequences=Displayed;
CC Note=No experimental confirmation available;
CC Name=2;
CC IsoId=Q93WV0-2; Sequences=VSP_007247, VSP_007248;
CC Note=No experimental confirmation available;
CC -!- SIMILARITY: Belongs to the WRKY group I family.
CC -!- SIMILARITY: Contains 2 WRKY domains.
CC -!- CAUTION: Ref.2 sequences differ from that shown due to erroneous
CC gene model prediction.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AF425837; AAL13050.1; -;
CC EMBL: AL078465; CAB43860.1; ALT_SEQ.
CC EMBL: AL161565; CAB79519.1; ALT_SEQ.
CC EMBL: AY045892; AAK76566.1; -;
CC EMBL: AY150436; AAN12978.1; -;
CC PIR: T08930; T08930.
CC InterPro: IPR003657; WRKY.

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; AF224699; AAK28309.1; -
DR EMBL; AB010697; BAB11168.1; -
DR EMBL; AL13652; CAB87266.1; -
DR EMBL; AY084692; AAM61254.1; -
DR EMBL; AK117545; BAC42206.1; -
DR EMBL; T48481; T48481.1; -
DR InterPro; IPR003657; WRKY.
DR Pfam; PF03106; WRKY; 2.
DR PROSITE; PS08111; WRKY; 2.
KW Transcription regulation; DNA-binding; Nuclear protein; Repeat.
FT DNA_BIND 111 176 WRKY 1.
FT DNA_BIND 228 293 WRKY 2.
FT CONFLICT 148 148 L -> P (IN REF. 5).
FT CONFLICT 165 165 I -> F (IN REF. 4).
FT CONFLICT 183 183 S -> P (IN REF. 1).
SQ SEQUENCE 309 AA; 34910 MW; B83AF11B93F3909E CRC64;

Query Match 29.08; Score 884; DB 1; Length 309;
Best Local Similarity 47.74; Pred. No. 2.6e-42;
Matches 198; Conservative 30; Mismatches 55; Indels 132; Gaps 13;
QY 58 VPKFKSTPPSLPLSPSPSSYFAIPGLSPAEILDSVLLNSSLPSPTTGAFVAQ 117
DB 10 VPKFKTATP-----SPULSPSPYFTWPGULTPADFLDSPLLFTSSNLSPTTGTPPAQ 64
QY 118 SPNWKSSGG---NQIVKBEEDSFNSFPQTRSGPPASSTATYQSSNVTVTQQPMSFQ 174
DB 65 SLNY--NNGLLIDKNEIKVEDTT-----PP-----LFLPSMVT---QPLPQL 102
QY 175 EATKQDNFSGKGMKMTENSMSQSPSPETASVQTNHSGFQSDYGNYPQSQTLERRSD 234
DB 103 DLFKSEIMSNK-----TSQ 117
QY 235 DGVNWKYGVKQVKGNSPRSYKCTYPNCTPKKVERSL-DGQITVIVYKTHNPKPO 293
DB 118 DGVNWKYGVKQVKGNSPRSYKCTYPNCTPKKVERSLVKGQMIIVYKGNHHPKPO 177
QY 294 NTRNNSNSSLAIHPSNSIRTEIPDSYATHGSGQMSAATPENSISIGDDDFEQSSQ 353
DB 178 STKSSSTAIA---AHQNS-----SNGD----- 197
QY 354 KCKSGGDEYDEPDARWKIEGNEGMSAPGSTVREPRVVQTTSDIILDDGYRWK 413
DB 198 -----GKDIDGETEAKRWKE-EN-----VKEPRVVQTTSDIILDDGYRWK 241
QY 414 YGQVVKGNPNRSPYKCTHGPCVKRVERASHDLRAVITTYEGKHNDVPAAR 468
DB 242 YGQVVKGNPNRSPYKCTTGCGFKRVERAFQDPKSVITTYEGRKHGQIPTPR 296

RESULT 5

WRK4 ARATH STANDARD; PRT; 514 AA.
ID WRK4 ARATH
AC QX190; Q93WN8; Q9LWGI;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Probable WRKY transcription factor 4 (WRKY DNA-binding protein 4).
GN WRKY4 OR ATG13960 OR F7A19.5 OR F16A14.18.
OS Arabidopsis thaliana (Mouse ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A., AND INDUCTION.
RX MEDLINE=21342502; PubMed=11449049;
RA Yu D., Chen C., Chen Z.;

RT RT
RT regulation of NPR1 Gene expression.";
RT Plant Cell 13:1527-1540 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia; TISSUE=Flower;
RA Ulker B., Kushnir S., Somssich I.E.;
RT "Arabidopsis thaliana transcription factor WRKY4.";
RN Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,
Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,
Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
Dunn P., Egu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,
Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,
Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziali A.,
Miltischer J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
Utterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;
RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
thaliana.";
RL Nature 408:816-820 (2000).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RT "RIKEN Arabidopsis full length cDNA clones (RAFLs) sequenced by the
SSP consortium (Salk/Stanford/PGEC).";
RN Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
RN [5]
RP TISSUE SPECIFICITY, AND INDUCTION.
RX MEDLINE=21580393; PubMed=1172756;
RA Robatzek S., Somssich I.E.;
RT "A new member of the Arabidopsis WRKY transcription factor family,
AtWRKY6, is associated with both senescence- and defence-related
processes.";
RL Plant J. 28:123-133 (2001).
CC -|- FUNCTION: Transcription factor. Interacts specifically with the W
box (5'-(T)TGAC(C/T)-3'), a frequently occurring elicitor-
responsive cis-acting element (By similarity).
CC -|- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -|- TISSUE SPECIFICITY: In young, mature and senescent leaves.
CC -|- INDUCTION: By salicylic acid and strongly during leaf senescence.
CC -|- SIMILARITY: Contains 2 WRKY domains.
CC -|- CAUTION: Ref.3 sequences differ from that shown due to erroneous
gene model prediction.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AF224703; AAK28313.1; -
CC EMBL; AF425835; AAL13048.1; -
CC EMBL; AC007576; AAD39282.1; ALT_SEQ.
CC EMBL; AC068197; AAF79402.1; ALT_SEQ.
CC EMBL; AY045676; AAK74034.1; -
CC EMBL; BT002629; AAO11545.1; -
CC PIR; G86272; G86272.
CC InterPro; IPR003657; WRKY.

DR Pfam; PF031106; WRKY, 2.
 DR PROSITE; PS50811; WRKY, 2.
 KW Transcription regulation; DNA-binding; Nuclear protein; Repeat.
 FT DNA_BIND 223 287 WRKY 1.
 FT DNA_BIND 403 468 WRKY 2.
 FT CONFLICT 112 147 AVLDLNCI (IN REF. 1).
 FT CONFLICT 112 147
 SQ SEQUENCE 514 AA; 55815 MW; 01010F8745C420C5 CRC64;
 Query Match 28.2%; Score 861.5; DB 1; Length 514;
 Best Local Similarity 38.3%; Pred. No. 8.1e-41;
 Matches 217; Conservative 62; Mismatches 139; Indels 149; Gaps 21;
 QY 31 SPSDLLASPLDNNKPPQGLS-----ERTGSGV-----PKEKSTPPSLPLSPPP 75
 DB 60 SFSQLLAGSSPATAAAAAATASYQRLGEGTSSSGDVPDRFKQNPRTGLMISQSQ 119
 QY 76 ISPSYFAIPGLSPAEILLSPVLLNSNLTSPSTTGAFVAQSPFNKSSGGNQIVKEE 135
 DB 120 -SP-SMTVPFGLSPAMLLDSPFLG---LSPVQGSY-----GWTQQAQAQ- 162
 QY 136 DKSPNSFQTRSGPPASSTATYQSSNVVQTQOPWFSQBATKQDNFSSGKGMKTNSS 195
 DB 163 -----VTAQAVQANANNQPTQTEP-----PPSQVQSPSSGQAQIPTSPAL 202
 QY 196 SMQSFSPELASVQTNHNSGFSQSDYGNYPQSQOTLSRSDGYNWRKYGQKOVKGSNP 255
 DB 203 PAQRETSQVTHIE-----HRSQPLNVDPKADDDGYNWRKYGQKOVKGSNP 250
 QY 256 YKCTYPCPTKKKVERSLDQITEIVYKGNHHPKQNTRR-NSSNSSSLAIPHSNISIR 314
 DB 251 YKCTNPGCPVKKVERSLDQVTEIYKGNHHPKQNTKRGKNDTANI---NGSSIN 307
 QY 315 TEIPDQSVATHGQMSQATPENSISIGDDPQSSQKC-----KSGGDEY----- 362
 DB 308 N-----NRGSELGASQPTNSNKTREQHEAVSQATTEHLSEASDGEVGNGET 359
 QY 363 -----DEDEPDAKWKLEGE-NEGMSAPGRTVREPVVQTTSDIILDGVRWRKYGQ 416
 DB 360 DVREKDNPEPKRRSTRVIRSEPAASHRTVTPEIRIIVTTSSEVLLDDGVRWRKYGQ 419
 QY 417 KVKGNPNRPSYKCTHPGCPVRKHVERASHDLRAVITVEGKINHDPVPAARGSGSHVN 476
 DB 420 KVKGNPNRPSYKCTHPGCPVRKHVERASHDLRAVITVEGKINHDLPAK-SSSH--- 475
 QY 477 RMPNPNASNTNTAATSVRLLPVTHQSDNSLQNSQAPPEGQSPFTLEMLQSPSGFSS 536
 DB 476 -----AAAAAQLRP-----DN-----RPG----- 489
 QY 537 GFGNPMQSYNQQLSDNVSSRTKEE 563
 DB 490 GLAN-----LNQOQQQOPVARLRUKEE 511
 RESULT 6
 ID WRKY_ARATH STANDARD; PRT; 687 AA.
 AC Q9FG77;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Probable WRKY transcription factor 2 (WRKY DNA-binding protein 2).
 GN WRKY2 OR AT5G56270 OR MKK23.1 OR K24C1.9.
 OS Arabidopsis thaliana (Mousse-eare cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
 ON NCBI_TaxID=3702;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA Ulker B., Kuehnir S., Somsich I.E.;
 FT "Arabidopsis thaliana transcription factor WRKY2.";

RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RX MEDLINE=21016721; PubMed=11130714;
 RA Tabata S., Kaneko T., Nakamura Y., Kotani H., Kato T., Asamizu E.,
 RA Miyajima N., Sasamoto S., Kimura T., Hosouchi T., Kawashima K.,
 RA Kohara M., Matsumoto M., Matsuura A., Muraki A., Nakayama S.,
 RA Nakazaki N., Naruo K., Okumura S., Shingo S., Takeuchi C., Wada T.,
 RA Watanabe A., Yamada M., Yasuda M., Sato S., de la Bastide M.,
 RA Huang E., Spiegel L., Gnoj L., O'Shaughnessy A., Preston R.,
 RA Haberman K., Murray J., Johnson D., Rohlfing T., Nelson J.,
 RA Stoneking T., Pepin K., Spieth J., Sekhon M., Armstrong J., Becker M.,
 RA Balter E., Cordum H., Cordes M., Courtney L., Courtney W., Dante M.,
 RA Du H., Edwards J., Fryman J., Haakensen B., Lamar E., Latreille P.,
 RA Leonard S., Meyer R., Mulvaney E., Ozersky P., Riley A., Stromatt C.,
 RA Wagner-McPherson C., Wollam A., Yeakam M., Bell M., Dedhia N.,
 RA Parnell L., Shah R., Rodriguez M., Hoon See L., Vil D., Baker J.,
 RA Kirchoff K., Toth K., King L., Bahret A., Miller B., Marra M.,
 RA Martensen R., McCombie W.R., Wilson R.K., Murphy G., Bancroft I.,
 RA Volckaert G., Wambutt R., Duesterhoeft A., Stiekema W., Pohl T.,
 RA Entian K.-D., Terry N., Hartley N., Bent E., Johnson S.,
 RA Langham S.-A., McCullagh B., Robben J., Grynoprez B., Zimmermann W.,
 RA Ramsperger U., Wedler H., Balke K., Wedler E., Peters S.,
 RA van Staveren M., Dirkee W., Moolman P., Klein Lankhorst R.,
 RA Weitzensagger T., Bothe G., Rose M., Hauf J., Berner S., Hempel S.,
 RA Feldpausch M., Lanberth S., Villarroel R., Gielen J., Ardiles W.,
 RA Bents O., Lemcke K., Kolesov G., Mayer K.F.X., Rudd S., Schoof H.,
 RA Schueller C., Zaccaria P., Mewes H.-W., Bevan M., Franz P.F.;
 RT "Sequence and analysis of chromosome 5 of the plant Arabidopsis
 thaliana.";
 RL Nature 408:823-826 (2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RA Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
 RT "RIKEN Arabidopsis full length cDNA clones (RAFLs) sequenced by the
 SSP consortium (Salk/Stanford/PGEC).";
 RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP TISSUE SPECIFICITY.
 RX MEDLINE=21580393; PubMed=11722756;
 RA Robatzek S., Somsich I.E.;
 RT "A new member of the Arabidopsis WRKY transcription factor family,
 AtWRKY6, is associated with both senescence- and defence-related
 processes.";
 RL Plant J. 28:123-133 (2001).
 CC -1- FUNCTION: Transcription factor. Interacts specifically with the W
 box (5'-(T)TGAC(C/T)-3'), a frequently occurring elicitor-
 responsive cis-acting element (By similarity).
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- TISSUE SPECIFICITY: Low expression in senescent leaves.
 CC -1- SIMILARITY: Belongs to the WRKY class I family.
 CC -1- SIMILARITY: Contains 2 WRKY domains.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See http://www.isb-sib.ch/announce/
 or send an email to license@isb-sib.ch).
 CC EMBL; AF418308; AL13039.1; -
 CC EMBL; AB026656; BAB08871.1; -
 CC EMBL; AY096493; AAM20132.1; -
 CC EMBL; AY123006; AAM67539.1; -
 CC InterPro: IPR003657; WRKY.
 CC Pfam; PF03106; WRKY, 2.
 DR PROSITE; PS50811; WRKY, 2.
 KW Transcription regulation; DNA-binding; Nuclear protein; Repeat.
 FT DOMAIN 437 446 POLY-GLY.
 FT DNA_BIND 267 331 WRKY 1.

```
FT DNA_BIND 481 546 WRKY 2.
SQ SEQUENCE 687 AA; 74561 MW; D47EAB1F80C6335F CRC64;

Query Match
Best Local Similarity 37.18; Pred. No. 3.8e-40; Length 687;
Matches 229; Conservative 63; Mismatches 146; Indels 180; Gaps 24;

QY 47 QGGLSE-----RTGSGVFKFKSTPPSPPLSPSPSSYFAI-----PGLSP 90
DB 72 RGLSERIAARAGNAPR-----LNTENIRNTWDFSIDNLSRSPCLTISSPGLSP 121
QY 91 AELLDSVPLNSSLNLSPTTGAFVAOSFNWKSNGN-----QQIVKE--DKSFSNFSQ 145
DB 122 ATLLSEPVFLSNPLAQSPTTG-----KFPFLPGVNGNALSEKAKDEFFDDIGASFSEH 176
QY 146 TRSGPPASSTATQSSNVTVTQOPWFSQBATQDNFSSGKGMMKTNSSSSMSQSFPEIA 205
DB 177 -----PVSRSSTP-----FQGTTEMMSVDYGNVNRSSSHQSAEVRPGSE 218
QY 206 SVQTNHNSGQSDYGNYPQSQTL-----SRRSD-----DGY 237
DB 219 NIESSNLYGIETD--NQNGQNKTSVTTNTSLETVDHQBEEBQRGRGDSMAGAPAEADGY 276
QY 238 NWRKYGQKVGKSNPRSYKCTYPNCPCTKKKVERSLEDGQITIVYKGTNHHKAPQNTNR 297
DB 277 NWRKYGQKVGKSEYPRSYKCTNPNCQVKKKVERSREGHITIIYKGAHNLKPPPNRR 336
QY 298 N-----SSNSSLAIPHNSIRTEIPDQSYAT 324
DB 337 SGMOVDTEQVEQQQQQDAAATVWSCNNTQQQGGNNENN-----EGSTTFEYGNQSGSI 393
QY 325 HG--SGQMDS-----AATPENSISIGDDDFEQSSQKCKS-----GGDEYDEDEPAKR 371
DB 394 QAQTGGQYSGDPVVVDVASTSNDDEDDRGTHGSVSLGYDGGGGGGGGGDESESKR 453
QY 372 WKIEGENEGMAGPSRTVRPRVVQTSIDILDGGRWKYKQKVKGNPNRPSYK 431
DB 454 RLEAFAAEMSG--STRAIREPRVVQTSIDILDGGRWKYKQKVKGNPNRPSYK 512
QY 432 THGPCVRKHEVRASHDLRAVITYEGKHNHDPAAAR-----GSGSHSVNRPMPNNASNHT 487
DB 513 TAGCTVRKHEVRASHDLKSVITYEGKHNHDPAAARNSSHGSGGSGN-----GNSG 565
QY 488 NTAATSRLLPVHQSDNSLQNRQSOAPPEG-----QSPFTLEMLQSPGSF-----533
DB 566 GSAAVS-----HHYHNGHSEPRGRDROVTTNNQSPFS-----RPFSPQPHL 609
QY 534 -----GFSGFGNPMQSYN 547
DB 610 GPPSGFS-FGLGQTGLVN 626

RESULT 7
WRKY_ARATH
ID WRKY_ARATH STANDARD; PRT; 513 AA.
AC Q9ZQ70;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 13-SEP-2003 (Rel. 42, Last annotation update)
DE Probable WRKY transcription factor 3 (WRKY DNA-binding protein 3).
GN WRKY3 OR AT2G03340 OR T4M8.23.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN SEQUENCE FROM N.A., AND INDUCTION.
RP MEDLINE=21342502; PubMed=11449049;
RX Yu D., Chen C., Chen Z.;
RA "Evidence for an important role of WRKY DNA binding proteins in the
RT regulation of NPR1 gene expression."
RL Plant Cell 13:1527-1540 (2001).
```

```

238 HRSQPNADPADGYNWRKYGQKVGSDFRPSYYKCTHPACPVKKVSKVSLDQGVTEI 297
282 VYKTHNHPQNTNRSSNSSSLAIPHNSIRTEIPDQYATHGSGQMSAATPENSII 341
298 IYKQHNHLPKRGNGNNGCKS-----SDIANQ-FQTSNLSLNKSKRDQTSQV 346
342 SIGDDDFEQSSQK-----CKSGGDEYDEPDKAEKWKTEGENEGMSAPGSRVREPRVVQ 397
347 TTTEQMSAASEVGNNAETSVGERHEDEPDKRNTETVRSEVPVASSHRTVTEPRIVQ 406
398 TTSDIDILDCGYRWRKYGQKVGKGNPNRPSYYKCTHPGCPVRKHVERASHDLRAVITYE 457
407 TTSEVDLLDGYRWRKYGQKVGKGNPYRPSYYKCTTDCGVRKHVERAATDPRAVVTYE 466
458 GKNHVDVPAARGSGSHSVNRPMPNNAHNTT 489
467 GKNHVDVPAARTS-SHQLR-----PNNQHN-TST 493

RESULT 8
WR34 ARATH
ID WR34 ARATH STANDARD; PRT; 568 AA.
AC O65590;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DE Probable WRKY transcription factor 34 (WRKY DNA-binding protein 34).
GN WRKY34 OR AT4G26440 OR M359.130.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
SEQUENCE FROM N.A.
STRAIN=cv. Columbia; TISSUE=Flower;
RA Ulker B., Kuehnir S., Somasich I.E.;
RT "Arabidopsis thaliana transcription factor WRKY34.";
RN [2]
SEQUENCE FROM N.A.
STRAIN=cv. Columbia; PubMed=10617198;
RX MEDLINE=20083488;
RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
RA Harris B., Ansoorge W., Brandt P., Grivell L.A., Rieger M.,
RA Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,
RA Kreis M., Delseny M., Puigdomenech P., Watson M., Schmidheini T.,
RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
RA Vos P., Hehseisel J., Zimmermann W., Wedler H., Ridley P.,
RA Langham S.-A., McCullagh B., Billam L., Robben J.,
RA Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,
RA Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
RA Weitzenecker T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
RA Holzer E., Brandt A., Peters S., van Staveren M., Dirckx W.,
RA Mooijman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
RA Berneser S., Hempel S., Feldpausch M., Lambert S., Van den Daele H.,
RA De Keyser A., Buysshaert C., Gielen J., Villarroel R., De Clercq R.,
RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McLay K., Mayes R.,
RA Pettett A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,
RA Borkova D., Bloeker H., Scharfe M., Grimm M., Loehnert T.-H.,
RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzi A.,
RA Neumann S., Argiriou A., Vitale D., Liquori R., Piravandi E.,
RA Massenot O., Quigley F., Clabaud G., Muendlein A., Felber R.,
RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
RA Chedof F., Cooke R., Berger C., Monfort A., Casacuberta E.,
RA Gibbons T., Weber N., Vandenbol M., Barges M., Terol J., Torres A.,
RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Bielke C.,
RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,

```

```

RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA Kramer J., Fulton L., Mardis E., Pante M., Pepin K., Hillier L.,
RA Nelson J., Spith J., Ryan E., Andrews S., Geisel C., Layman D.,
RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Tili S.,
RA Grant S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
RA Chen E., Maria M., Martienssen R., McCombie W.R.;
RT "Sequence and analysis of chromosome 4 of the plant Arabidopsis
thaliana.";
RL Nature 402:769-777(1999);
CC -!- FUNCTION: Transcription factor. Interacts specifically with the W
CC box (5'-(T)TGAC(C/T)-3'), a frequently occurring elicitor-
CC responsive cis-acting element (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- SIMILARITY: Belongs to the WRKY class I family.
CC -!- SIMILARITY: Contains 2 WRKY domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.emb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; AY052649; AAL11010.1; -
CC EMBL; AL022223; CAB18226.1; -
CC EMBL; AL161565; CAB79499.1; -
CC PIR; T05060;
CC InterPro; IPR003657; WRKY.
CC Pfam; PF03106; WRKY; 2.
CC PROSITE; PS0811; WRKY; 2.
KW Transcription regulation; DNA-binding; Nuclear protein; Repeat.
FT DNA_BIND 172 236 WRKY 1.
FT DNA_BIND 366 431 WRKY 2.
SQ SEQUENCE 568 AA; 62166 MW; 31F350AA46341F7B CRC64;

Query Match 24.3%; Score 742; DB 1; Length 568;
Best Local Similarity 37.3%; Pred. No. 3.6e-34;
Matches 193; Conservative 55; Mismatches 119; Indels 150; Gaps 16;

QY 86 PGLSPAEILDSPVLLNSSNLPSPPTTGAFVAQSFNWKSSGGNQIQIVKEEDKSFNFSQ 145
DB 88 PGLSPATLLESVPVL--SNPLLSPTTGKL-----SSVPSDKAKAELFDDITTSIAFQ 137
QY 146 TRGSPASSTATTQSSNVTVQTQPPNSFQATKQDNFSSGKGMKMTENSSMSQSPFSEA 205
DB 138 TISGSGLDPT-----NIALEPDDSDQYBERQ-----LGLGLG-----DSMACCAP--- 176
QY 206 SVQTNHNSGQSDYGNVPQSTLSRRSDGYNWRKYGQKVGKGNPRSYKCTVPNCP 265
DB 177 -----ADGYNWRKYGQKVGKSEYPRSYKCTHPNCE 209
QY 266 TKKKVRSRLDQGITTEIVYKGTHTNHPKQNTNRSSNSSSLA----- 306
DB 210 AKKKVRSRREGHIIETITGDIHSPKPPNRRRRSGISGSGTGQDMQIDATEYEGFAGTNN 269
QY 307 IPHNSIRTEIPQSYA-----THGSGQMSAATPENSISIGDDDFEQSSQKCKSG 358
DB 270 IEWTSFVSALEYGSHSGSMQVNGTHQFGYGA---DAILYRDENEEDRTSHMSVSL 325
QY 359 GDEYDEDEPDKAEKWKTEG---ENEGMSAPGSRVREPRVVVQTTSIDILDDGYRWRKYG 415
DB 326 TYGEVEEESKARKLEAVATETSG-----STRASRPRVVVQTTSIDILDDGYRWRKYG 381
QY 416 QKVVGKGNPNRPSYYKCTHPGCPVRKHVERASHDLRAVITYEGKHNHVDVPAAR-----GS 470

```

Db 382 QKVGKGNPRSYKCTANGCTVTKHVERASDDFKSVLLTYIGKTHVPAARNSSHVGA 441
QY 471 GS-----HSVNRPMNPNASHTNTAATSVRLLPVTHOSDNLQNRQOAPP 516
Db 442 GSGGTGGLSLATQTHNHNHYPHRSRSEGLATA-----NSLFDQSH--- 485

QY 517 EGQSPFTLEMLQSPGSGFGFGNPMQSYVNOQLSD 553
Db 486 -----LRHTGFSV-----YIGQSELS 503

RESULT 9

WR58 ARATH STANDARD; PRT; 423 AA.
ID WR58 ARATH STANDARD; PRT; 423 AA.
AC Q93WU7; O9MAC2;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 15-SEP-2003 (Rel. 42, Last annotation update)
DE Probable WRKY transcription factor 58 (WRKY DNA-binding protein 58).
GN WRKY58 OR AT3G01080 OR T4P13.24.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
ON NCBI_TaxID=3702;
RX MEDLINE=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia; TISSUE=Flower;
RA Kushnir S., Ulker B., Somsich I.E.;
RT "Arabidopsis thaliana transcription factor WRKY58."
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016720; PubMed=11130713;
RA Salancubut M., Lemcke K., Rieger M., Ansoerge W., Unseled M.,
RA Fartmann B., Valle G., Bloeker H., Perez-Alonso M., Obermaier B.,
RA Delseny M., Boutry M., Grivell L.A., Mache R., Puigdomenech P.,
RA De Simone V., Choine N., Artiguenave F., Robert C., Brottier P.,
RA Wincker P., Catalicio L., Weissbach J., Saurin W., Quetier F.,
RA Schaefer M., Mueller-Auer S., Gabel C., Fuchs M., Benes V.,
RA Wurmbach E., Drzonek H., Floss H., Jordan N., Bangert S.,
RA Wiedelmann R., Kranz H., Holland R., Brandt P., Nyakatura G.,
RA Vezzi A., D'Angelo M., Pallavicini A., Toppo S., Simonati B.,
RA Conrad A., Hornischer K., Kauer G., Loehner T.-H., Nordsiek G.,
RA Reichelt J., Scharfe M., Schoen O., Bargues M., Terol J., Climent J.,
RA Navarro P., Collado C., Perez-Perez A., Ottenwaelder B., Duchemin D.,
RA Cooke R., Laudie M., Berger-Llauro C., Purnelle B., Masuy D.,
RA de Haan M., Maarse A.C., Alcaraz J.-P., Cottet A., Casacuberta E.,
RA Monfort A., Argiriou A., Flores M., Liguori R., Vitale D.,
RA Mannheim G., Haase D., Schoof H., Rudd S., Zaccaria P., Mewes H.-W.,
RA Mayer K.F.X., Kaul S., Town C.D., Koo H.L., Tallon L.J., Jenkins J.,
RA Rooney T., Rizzo M., Walts A., Utterback T., Fujii C.Y., Shea T.P.,
RA Creasy T.H., Haas B., Maiti R., Wu D., Peterson J., Van Aken S.,
RA Pai G., Miltischer J., Sellers P., Gill J.E., Feldblyum T.V.,
RA Preuss D., Lin X., Nierman W.C., Salzberg S.L., White O., Venter J.C.,
RA Sasamoto S., Kimura T., Idesawa K., Kawashima K., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakayama S., Nakazaki N., Shinpo S., Takeuchi C., Wada T.,
RA Watanabe A., Yamada M., Yasuda M., Tabata S.;
RT "Sequence and analysis of chromosome 3 of the plant Arabidopsis thaliana."
RL Nature 408:820-822 (2000).
CC -1- FUNCTION: Transcription factor. Interacts specifically with the W box (5'-(T)TGAC(C/T)-3'), a frequently occurring elicitor-responsive cis-acting element (By similarity).
CC -1- SUBCELLULAR LOCATION: Nuclear (Potential).
CC -1- SIMILARITY: Contains 2 WRKY domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
CC -----

DR EMBL; AF426254; AAL29431.1; -;
DR EMBL; AC008261; AAF26166.1; -;
DR InterPro; IPR003657; WRKY.
DR Pfam; PF03106; WRKY; 2.
DR PROSITE; PS50811; WRKY; 2.
KW Transcription regulation; DNA-binding; Nuclear protein; Repeat.
FT DNA BIND 161 225 WRKY 1.
FT DNA BIND 300 365 WRKY 2.
FT DOMAIN 150 154 POLY-ASN.
FT DOMAIN 257 270 POLY-ASP.
FT CONFLICT 81 81 L -> LNTGMSVSPGGGRSTAGMPAGGFMFTIPSGFSPS
FT FT SLL (IN REF. 2).
SQ SEQUENCE 423 AA; 47121 MW; EF6C2F1BF3F16B15 CRC64;

Query Match 24.2%; Score 738; DB 1; Length 423;
Best Local Similarity 37.8%; Pred. No. 4.2e-34;
Matches 182; Conservative 52; Mismatches 143; Indels 104; Gaps 14;
QY 65 PPSLPLSPPPSPSYF-----AIPGLSPAELLDSPVLLNSNILLPSPTTGAFAVAQ 117
Db 21 PTTVPHRPPAETAAAYFFGGGGLSLSPG--PLSFVSLFVDNFPDVL----- 67

QY 118 SFNWKSSSGNQIIVKEEDKSPNSFQTRSGPPASSTAT-----YQSNVTVTQOPWS 172
Db 68 -----TPDNQRTT-----SFTQLTSPMFFPOSSAHTGFIQPRQSQ--PQQRPD 113

QY 173 FQEATQDNFSGKGMKMTENSSMSQSFPEIASVQTNHSGFQSDYGNYPQSTLSRR 232
Db 114 FPHMPPSTSVAVHGQSLDVS-----QVDQARNHNNPNNNNRSYVNVVDKP 165

QY 233 SDGYNWRKYQKQKNGSENPRSYKCTPKKKKVERSLEDQITEIVYKTHNPKP 292
Db 166 ADDGYNWRKYQKPIKGCYPRSYKCTHVCNCPKVVKVERSLEDQITQIYKQHDHERP 225

QY 293 QNTRNSNSSSLAIPHSNIRTEIPQSYATHGSCQMSAATPENSISIGDDDFEQSS 352
Db 226 QNRGGGGRDS-----TEV-----GGAGQMMESSDDSGYRKHDDDDDD-- 264

QY 353 QKCKSGGDRVDEDPDAKRWKTEGENEGMSAPGSRVVRVVTQTSIDILDDGVWR 412
Db 265 -----DEDDDLPAKRRIDG-----VSTHRTVTEPKIIVQTKSEVDDLDGVRW 312

QY 413 KYGQKVVGNPNRPSYKCTHFCPCVYRKHVERASHDLRAVITYEGKHNHDPVARGSGS 472
Db 313 KYGQKVVGNPNRPSYKCTHFCPCVYRKHVERASTDAKAVITYEGKHNHDPVARGSGS 370

QY 473 HSNVRPMNPNASHTNTAATSVRLLPVTH-----QSDNSLQNRQ--QAPPEQSQFTLE 525
Db 371 -----TAAATAAAGVPSDHRMSGNNMQHSMFGNNNNNTQSPVLLR 415

QY 526 M 526

Db 416 L 416

RESULT 10

WR44 ARATH STANDARD; PRT; 429 AA.
ID WR44 ARATH STANDARD; PRT; 429 AA.
AC Q9ZUU0;
DT 15-SEP-2003 (Rel. 42, Created)
DT 15-SEP-2003 (Rel. 42, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE WRKY transcription factor 44 (WRKY DNA-binding protein 44)
DE (TRANSPARENT TESTA GLABRA 2).
GN WRKY44 OR TTG2 OR AT2G37260 OR P3G5.5.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

RA Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,
 RA Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
 RA Venter J.C.;
 RT "Sequence and analysis of chromosome 2 of the plant *Arabidopsis*
 RT *thaliana*.";
 RL Nature 402:761-768 (1999).
 CC -1- FUNCTION: Transcription factor. Binds to a 5'-CGTGGACGAG-3'
 CC consensus core sequence which contains a w box, a frequently
 CC occurring elicitor-responsive cis-acting element.
 CC -1- COFACTOR: Requires metal-ions (probably zinc) for its binding
 CC activity.
 CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -1- ALTERNATIVE PRODUCTS:
 CC Event=Alternative splicing; Named isoforms=2;
 CC Name=1;
 CC IsoId=Q9SI37-1; Sequence=Displayed;
 CC Name=2;
 CC IsoId=Q9SI37-2; Sequence=VSP_007124;
 CC -1- TISSUE SPECIFICITY: Expressed to similar levels in root and
 CC flower, to a somewhat lower level in stem and to low levels in
 CC leaf and siliques.
 CC -1- MISCELLANEOUS: Binding to target DNA is mediated mainly by the C-
 CC terminal WRKY domain, while part of the activation domain is
 CC located between positions 210 and 285.
 CC -1- SIMILARITY: Belongs to the WRKY class I family.
 CC -1- SIMILARITY: Contains 2 WRKY domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; X92976; CAA63554.1; -;
 CC EMBL; AF442389; AAL35282.1; -;
 CC EMBL; AF442390; AAL35283.1; -;
 CC EMBL; AC007211; AAD25579.1; -;
 CC EMBL; AC006955; AAM15341.1; -;
 CC PIR; F84462; F84462.
 CC InterPro; IPR003657; WRKY.
 CC Pfam; PF03106; WRKY; 2.
 CC PROSITE; PS50811; WRKY; 2.
 CC Transcription regulation; DNA-binding; Zinc; Nuclear protein; Repeat;
 KW Alternative splicing; Activator
 FT DNA_BIND 105 169 WRKY 1.
 FT DNA_BIND 301 366 WRKY 2.
 FT VARSPPLIC 191 214 Missing (in isoform 2).
 FT /FTID=VSP_007124.
 FT SQ SEQUENCE 487 AA; 54010 MW; 61712DBD16897C38 CRC64;
 Query Match 18.2%; Score 556.5; DB 1; Length 487;
 Best Local Similarity 35.0%; Pred. No. 5.1e-24;
 Matches 131; Conservative 50; Mismatches 98; Indels 95; Gaps 10;
 QY 234 DDGNNWRKYGKQVKGSGENSRYSYKCTYPNCPKTKKVSRLDQGIETIVYKGTNHPKP- 292
 DB 111 EDGNNWRKYGKLVKGNFEVRSYRCTHPNCKAKKQLERSAGGVVDTVYFGSHDPKPL 170
 QY 293 -----QNTRRNSNSS-----LAI-----PHSNSIRTEIPDQS 321
 DB 171 AGAVPINQDKRSVFTAVSGEQRIDIVSLIYKLCIVSDYIMFVKTSGSSVQLRQTEP 230
 QY 322 YATHGSGQMSDAATPENSISIGDDDFEQSQKCKSGKGDYDEB--EPDAKRWKIEGENE 379
 DB 231 PKIHGGLHV-----SVIPPADDVKTIDISQSSRITGDNTHKDYNSPTAKRRKKGNI 282
 QY 380 GMSAPGSRVREPRVVVTTSDILDGGRWKYGGKVGKGNPNRPSYKCTHPGCPVR 439
 DB 283 --LSPVSTNDGRIVVHTTFLDINVDGRWKYGGKVGKGNPNRPSYKCTHPGCPVK 340
 QY 440 KHVERASHDLRAVITTYEGKHNDVPAARGSGSHSVNRPMPNPNASHTNTAATSVRLLPV 499

Db 341 KHVERSSHDTKLLITTYEGKHNDMP-----PGRVVTHNN----- 375
 QY 500 IHQSDNSLONRSQAPPEGQSPTLEMLQSPGSGFGSGFNPMQSVVNOQLSDNVFSSR 559
 DB 376 --MLDSEVDDKEGDANKTFQS-----STLQSTKQHQHVEDHL----- 410
 QY 560 TKPEPRDDMFLESL 573
 DB 411 -RKKTGTNGFEKSL 423
 RESULT 12
 WR19 ARATH
 ID WR19 ARATH STANDARD; PRT; 1895 AA.
 AC Q9SZ67;
 DT 15-SEP-2003 (Rel. 42, Created)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Probable WRKY transcription factor 19 (WRKY DNA-binding protein 19).
 GN WRKY19 OR AT4G12020 OR F16J13.90.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
 ON NCBI_TaxID=3702;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Columbia;
 RC MEDLINE=20083488; PubMed=10617198;
 RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
 RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
 RA Harris B., Anseorge W., Brandt P., Grivell L.A., Rieger M.,
 RA Weichselgartner M., de Simone V., Obermaier B., Maché R., Mueller M.,
 RA Kreis M., Delsen W., Putgdomenech P., Watson M., Schmidtheini T.,
 RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
 RA Vos P., Hoheisel J., Zimmermann W., Wedler H., Ridley P.,
 RA Langham S.-A., McCullagh B., Bilham L., Robben J.,
 RA Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,
 RA Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
 RA Weitzenecker T., Bothe G., Ramepger U., Hilbert H., Braun M.,
 RA Holzer E., Brandt A., Peters S., van Staveren M., Dirkee W.,
 RA Moolijman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
 RA Berneriser S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
 RA De Keyser A., Buyschaert C., Gielen J., Villarroel R., De Clercq R.,
 RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
 RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McLeay K., Mayes R.,
 RA Pettett A., Rajadream M.A., Lyne M., Benes V., Rechmann S.,
 RA Borkova D., Bloecker H., Scharfe M., Grimm M., Loehner T.-H.,
 RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
 RA Gabel C., Fuchs M., Partmann B., Grandrath K., Dauner D., Herzl A.,
 RA Neumann S., Argiriou A., Vitale D., Liguori R., Piravandi E.,
 RA Massenot O., Quigley F., Clabaud G., Muendlein A., Felber R.,
 RA Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
 RA Chedor F., Cooke R., Berger C., Monfort A., Casacuberta E.,
 RA Gibbons T., Weber N., Vandenbol M., Bagues M., Terol J., Torres A.,
 RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
 RA Heijnen L., Schwarz S., Scholler P., Heber S., Francis P., Biele C.,
 RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
 RA Zaccaria P., Devian M., Wilson R.K., de la Bastide M., Habermann K.,
 RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
 RA Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
 RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
 RA Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
 RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
 RA Kramer J., Fulton L., Mardis E., Dente M., Pepin K., Hillier L.,
 RA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
 RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
 RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
 RA Ma P., Zhong J., Preston R., Vill D., Shekher M., Matero A., Shah R.,
 RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
 RA Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
 RA Chen E., Marra M., Martienssen R., McCombie W.R.;

RA Weitzenecker T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
RA Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,
RA Moeljan P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
RA Barneiser S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
RA De Keyser A., Buyschaert C., Gielen J., Villarroel R., De Clercq R.,
RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Alen S.,
RA Clark L., Doggett J., Hall S., Kay M., Lennard N., McLaughlin K., Mayes R.,
RA Pettett A., Rajadream M.A., Lyne M., Benes V., Rechmann S.,
RA Borkova D., Bloeker H., Scharfe M., Grimm M., Loeckert T.-H.,
RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzl A.,
RA Neumann S., Argirou A., Vitale D., Ligouri R., Piravandi E.,
RA Massenet O., Quigley F., Clabaud G., Muehlen A., Feilner R.,
RA Schnabl S., Hiller R., Schmidt W., Iechary A., Aubourg S.,
RA Cheford F., Cooke R., Berger C., Monfort A., Casacuberta E.,
RA Gibbons T., Weber N., Vandenberg M., Bagues M., Terol J., Torres A.,
RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
RA Heijnen L., Schwarz S., Scholler P., Heber S., Frances P., Bielek C.,
RA Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
RA Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
RA Sakon M., Murray J., Sheet J., Cordes M., Abu-Threideh J.,
RA Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
RA Latrelle P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.,
RA Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
RA Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
RA Antoniou B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
RA Ma P., Zhong J., Preston R., Vil D., Shekher M., Maturo A., Shah R.,
RA Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
RA Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
RA Chen E., Marra M., Martienssen R., McCombie W.R.,
RA "Sequence and analysis of chromosome 4 of the plant Arabidopsis
RA thaliana";
RA Nature 402:769-777(1999).
RN [2]
RN SEQUENCE FROM N.A.
RP STRAIN=cv. Columbia; TISSUE=Flower;
RC Shinozaki K., Davis R.W., Ecker J.R., Theologis A.;
RA "RIKEN Arabidopsis full length cDNA clones (RAFLs) sequenced by the
RT SFP consortium (Salk/Stanford/PGECC).";
RL Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: Transcription factor. Interacts specifically with the W
CC box (5'-(T)TGAC(C/T)-3'), a frequently occurring elicitor-
CC responsive cis-acting element (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- SIMILARITY: Belongs to the WRKY class I family.
CC -!- CAUTION: Ref.1 sequences differ from that shown due to erroneous
CC gene model prediction. AT4G30930 and AT4G30935 were originally
CC fused into a single gene.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AL022198; CAA18200.1; ALT_SEQ.
CC EMBL; AL161578; CAB79811.1; ALT_SEQ.
CC EMBL; BT004086; AAO42113.1; -.
CC PROSITE; PSS0811; WRKY; 2.
CC Transcription regulation; DNA-binding; Nuclear protein; Repeat.
CC DNA BIND 162 226 WRKY 1.
CC FT DNA BIND 325 390 WRKY 2.
CC SQ SEQUENCE 466 AA; 51480 MW; A4CC4E13B5262DAF CRC64;
CC
CC Query Match 15.8%; Score 482; DB 1; Length 466;
CC Best Local Similarity 39.3%; Pred. No. 6.2e-20;
CC Matches 118; Conservative 39; Mismatches 111; Indels 32; Gaps 10;

QY 204 IASVOTNHSNGFQSDYGN---YPPQSQTLSRSDGYNWRKYQKQKGVSEPRSYKCT 260
Db 138 LVSVPKQKQSDSPVNVRLSTVTPVPTPAR---DGYNWKYQKQKGVSEPRSYRCT 194
QY 261 YNCPYKPKKVERSLD-GQTEIVYKGTNHPKQPNTRNNSSSSLAIPHSINSIRTEIPD 319
Db 195 YTEC-CAKKIECSNDGNGVVEIKGLHTHEPRKT-----SFSPEIRVTTAIRPVSED 248
QY 320 QSVATH---GSGQMDSATPENSSISIGDDDFEQSSQKCKSGGDEYDEPDKEWKIE 375
Db 249 DTVEELSIIVPGSDPSASTKEYICES---QTLVDRKRHCENEAVE---EPEPKR-RLK 300
QY 376 GNEGMSAPGSRTRVPRVVVQVTTSDIDLDGYRWRKYQKQKGVSEPRSYKCTHPG 435
Db 301 KONSQSSDSVSPKPKKFKVHAAGDVIGDGYRWRKYQKQKGVSEPRSYRCTSAG 360
QY 436 CPVRKIVERASHDLRAVITYYEGKHNHVPAAAGSGSHSNRPMNPNNASHNTTAATSVR 495
Db 361 CPVRKHETAVENTKAVIITYYKGVHNDHMPVK-----KRHGPSPSSMLVAAAPTSMR 413

RESULT 14

WR10_0RATH

ID WR10_0RATH STANDARD; PRT; 506 AA.

AC 09LQ05; Q8VWQ3;

DT 15-SEP-2003 (Rel. 42, Created)

DT 15-SEP-2003 (Rel. 42, Last sequence update)

DT 15-SEP-2003 (Rel. 42, Last annotation update)

DE Probable WRKY transcription factor 10 (WRKY DNA-binding protein 10).

OS WRKY10 OR AT1G55600 OR F20N2.3.

GN Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.

OX NCBI_TaxID=3702;

RN [1]

RN SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia; TISSUE=Flower;

RA Ulker B., Kushnir S., Sonnsich I.E.;

RL "Arabidopsis thaliana transcription factor WRKY10.";

RN Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.

RN [2]

RN SEQUENCE FROM N.A.

RC STRAIN=cv. Columbia;

RX MEDLINE=21016719; PubMed=11130712;

RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,

RA White O., Alonso J., Altafi H., Araujo R., Bowman C.L., Brooks S.Y.,

RA Buehler E., Chan A., Chao Q., Chen H., Cheuk R.F., Chin C.W.,

RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,

RA Dunn P., Etgu P., Feldblyum T.V., Feng J.-D., Fong B., Fujii C.Y.,

RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Huizar L.,

RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,

RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,

RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,

RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,

RA Miltschker J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,

RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.M.,

RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,

RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,

RA Uterback T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,

RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.;

RT "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RT thaliana";
RL Nature 408:816-820(2000).
CC -!- FUNCTION: Transcription factor. Interacts specifically with the W
CC box (5'-(T)TGAC(C/T)-3'), a frequently occurring elicitor-
CC responsive cis-acting element (By similarity).
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- SIMILARITY: Belongs to the WRKY class I family.
CC -!- CAUTION: Ref.1 sequence differs from that shown due to erroneous
CC gene model prediction.

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: January 20, 2004, 16:16:52 ; Search time 40 Seconds
(without alignments)

3709.506 Million cell updates/sec

Title: US-09-890-811B-10

Perfect score: 3050

Sequence: 1 MASSSSGLDTSASANSFTNF.....FSSRTKEPRDDMFLESLIC 575

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1822	59.7	549	10	Q40090
2	1389.5	45.6	559	10	Q941B6
3	1379.5	45.2	515	10	O81639
4	1342	44.0	514	10	Q9SQ04
5	1331	43.6	514	10	Q40827
6	1278	41.9	477	10	Q9XJ26
7	1222	40.1	353	10	Q9SXP5
8	1201.5	39.4	378	10	Q941B4
9	1139.5	37.4	512	10	Q8S8P5
10	1116.5	36.6	501	10	Q9XFF1
11	1081	35.4	357	10	Q94D89
12	1032.5	33.9	402	10	Q38704
13	927	30.4	727	10	Q941B3
14	924	30.3	739	10	Q9ARCF
15	922	30.2	739	10	Q9ARD0
16	894.5	29.3	387	10	Q94AT4

17	894	29.3	309	10	Q8GYK8	Q8gyk8 arabidopsis
18	852	27.9	687	10	Q9FG77	Q9fg77 arabidopsis
19	843	27.6	571	10	Q9LMG1	Q9lmg1 arabidopsis
20	840	27.5	528	10	Q9ZPL6	Q9zpl6 nicotiana t
21	816.5	26.8	485	10	Q8H1E9	Q8h1e9 arabidopsis
22	812.5	26.6	485	10	Q94AP6	Q94ap6 arabidopsis
23	799	26.2	490	10	Q9SSX9	Q9ssx9 nicotiana t
24	778.5	25.5	509	10	Q93658	Q93658 cucumis sat
25	759.5	24.9	488	10	Q8W1M6	Q8w1m6 retama raet
26	742	24.3	568	10	Q65590	Q65590 arabidopsis
27	680.5	22.3	415	10	Q8SE29	Q8se29 oryza sativ
28	677	22.2	369	10	Q94IB5	Q94ib5 nicotiana t
29	613.5	20.1	439	10	Q9AUW5	Q9auw5 oryza sativ
30	609	20.0	591	10	Q947G7	Q947g7 solanum dul
31	574.5	18.8	463	10	Q43388	Q43388 arabidopsis
32	565.5	18.5	296	10	Q40829	Q40829 petroselinu
33	561.5	18.4	349	10	Q9ZU00	Q9zu00 arabidopsis
34	556.5	18.2	487	10	Q9SI37	Q9si37 arabidopsis
35	555	18.2	454	10	Q9ZPL7	Q9zpl7 nicotiana t
36	548	18.0	482	10	Q8H6W1	Q8h6w1 oryza sativ
37	520	17.0	1895	10	Q9SZ67	Q9sz67 arabidopsis
38	488.5	16.0	782	10	Q65556	Q65556 arabidopsis
39	390	12.8	513	10	Q9ARF4	Q9arf4 capsella ru
40	386	12.7	506	10	Q9LQ05	Q9lq05 arabidopsis
41	378.5	12.4	485	10	Q8VWQ3	Q8vwq3 arabidopsis
42	337	11.0	565	10	Q94E13	Q94e13 oryza sativ
43	329	10.8	379	10	Q9FE35	Q9fe35 oryza sativ
44	318	10.4	191	10	Q8S8S5	Q8s8s5 arabidopsis
45	318	10.4	218	10	Q93WY4	Q93wy4 arabidopsis

ALIGNMENTS

RESULT 1

Q40090	PRELIMINARY;	PRT;	549 AA.
ID	Q40090		
AC	Q40090;		
DT	01-NOV-1996 (TrEMBLrel. 01, Created)		
DT	01-NOV-1996 (TrEMBLrel. 01, Last sequence update)		
DT	01-OCT-2002 (TrEMBLrel. 22, Last annotation update)		
DE	SPF1 protein.		
OS	Ipomoea batatas (Sweet potato) (Batatae).		
OC	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;		
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;		
OC	Asteridae; lamids; Solanales; Convolvulaceae; ipomoea.		
OX	NCBI_TaxID=4120;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=Kokei No.14; TISSUE=Petiolo;		
RX	MEDLINE=95058910; PubMed=7969025;		
RA	Ishiguro S., Nakamura K.;		
RT	"Characterization of a cDNA encoding a novel DNA-binding protein, SPF1, that recognizes SP8 sequences in the 5'-upstream regions of genes coding for sporamin and beta-amylase from sweet potato."		
RL	Mol. Gen. Genet. 244:563-571(1994).		
DR	EMBL: D30038; BAA06278.1; --		
DR	TRANSFAC; T03975; --		
DR	InterPro; IPR003657; WRKY.		
DR	Pfam; PF03106; WRKY; 2.		
DR	PROSITE; PS0811; WRKY; 2.		
KW	DNA-binding.		
SQ	SEQUENCE 549 AA; 59695 MW; 5EC11DBCFF96FA3A CRC64;		
Query Match	59.7%;	Score 1822;	DB 10; Length 549;
Best Local Similarity	62.0%;	Pred. No. 4.4e-120;	
Matches 369;	Conservative 66;	Mismatches 92;	Indels 68; Gaps 14;
QY	1 MASSSSGLDTSASANSFTNFTHFTFTSPFDLLASPLDNNKPPQGGI----SRTGS 56		
DB	1 MAASSSTIDAPTASS---FSFTASSFMSSFTDLLASDAYSGSVSRGLDIAGRTGS 57		
QY	57 GVPKFKSTPPSLPLSPPIPSPSYFAIPGLSPAEILLOSPVLLNSSNLPSTTTGAFVA 116		

```

Db 58 GVFKFKSLPPSLSPSPSPSYFAFPGLSPSELLDSPVLLSSNLLPSTGTGTPA 117
Qy 117 QSNFWKSSGNGNQIVKEEDKGFNSFQTRSGPPASSTATYQSSNVTVOTQOPWMSQEA 176
Db 118 QTNWKNKSNASQEDVKEEKGYPDFSFQTN-----ASMTLYEDSK----- 160
Qy 177 TKQDNFSSGKGMKMTENSNSQSPSPSIASVQTNHSN--GFQSDYGNV--PQSQTL--SR 231
Db 161 -----RKDELNSLQSLPPVTTTQMSQQNGSGSYEYNNQCCPPSQTLUREQR 207
Qy 232 RSDDGYNWRKYQKQVKGSENPSPSYKCTYPCNCTKKKVERSILDGQITEIVYKGTNHPK 291
Db 208 RSDDGYNWRKYQKQVKGSENPSPSYKCTYPCNCTKKKVERALDGQITEIVYKGAHHPK 267
Qy 292 PONTNRNS-----SNSSLAIPHNSIRTEIPDQSYATHGSGQMDSAATPENSISIGDD 347
Db 268 PQSTRSSSSSTASSASTLAAQSYNAPASDVDPQSYWNGNGQMDSVATPENSISVGDDE 327
Qy 348 FEQSSQKCKSGGDEYDEDEPAKWKTEGENEGMSAPGSRVREPRVVVQTTSDIDLDD 407
Db 328 FEQSSQKCKSGGDEYDEDEPAKWKTEGENEGMSAPGSRVREPRVVVQTTSDIDLDD 387
Qy 408 GYRWKTYGQKQVKGSENPSPSYKCTYPCNCTKKKVERASHDLRAVITYYEGKKNHVPAA 467
Db 388 GYRWKTYGQKQVKGSENPSPSYKCTYPCNCTKKKVERASHDIRSVITYYEGKKNHVPAA 447
Qy 468 RGSQSHSVNRPMPNPNASNHTNTAATSVRLLPVTHQSDNSLQNG-----RQAPPEQ 519
Db 448 RGSQSHGLNR-----GANPNNAAMAMAIRP-----STMSLQSNYPIPISTRPWQGEQ 498
Qy 520 SPFTLEMLQSPGSGFGNGPMQSYUNQOQLSDNVFSSRTKEPRDDMLESL 574
Db 499 APY--EMLQSGGGYSGFGNPMNAYAN--QIQDNAP--SRAKSEPRDDLFLDTLL 548

```

RESULT 2

```

Q941B6 PRELIMINARY; PRT; 559 AA.
ID Q941B6
AC Q941B6;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE WRKY DNA-binding protein.
GN WRKY-6.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamiids; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Petit Havana SRI;
RA Maeo K., Hayashi S., Kojima-Suzuki H., Morikami A., Nakamura K.;
RT "Role of conserved residues of the WRKY domain in the DNA-binding
RT activity of tobacco WRKY family proteins.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB063573; BAB61053.1; -.
DR InterPro; IPR003657; WRKY.
DR Pfam; PF03106; WRKY; 2.
DR PROSITE; PS50811; WRKY; 2.
KW DNA-binding
SQ SEQUENCE 559 AA; 62288 MW; A0DD3549C811B5AE CRC64;

```

```

Query Match 45.6%; Score 1389.5; DB 10; Length 559;
Best Local Similarity 51.5%; Pred. No. 1.2e-89;
Matches 316; Conservative 70; Mismatches 123; Indels 105; Gaps 23;

Qy 1 MASSSGSLDTSASANSST-NFTFTSTHPTMTYFSDLLASPLDNNKPPQGG----- 49
Db 1 MASSGGNMT--FMNSTSYFS-----SFSDDLSDNNNNNNNNNSNMQEKNS 51
Qy 50 -----LSERTGSGVPKFKSTPPPSLPL--SPPISSPSYFAIPGLSPABELLSPV 98

```

```

Db 52 LNWGPDQRMHQKNDDEVKFKSPFPCSLPMTSSSPASPSSYLAPPSLSPSVLLDSPV 111
Qy 99 LNSSNLLSPPTGAFVAQSPFWKSSGNGNQIVKEEDKGFNSFQTRSGPPASSTATY 158
Db 112 LFNNSNLLSPPTGSP-----GNLN--SKEDDSRIIDFSFQSR-----ATSSMF 155
Qy 159 QSN-----VTVOTQOPWFSFOBATKQDNFSSGKGMKMTENSNSMQSPSPSIASVQTN 210
Db 156 QSSAPRNSLEDLWTRQ-----QHANOQNEFSTAKTTGVKSEVPIQSFQSE--KMQSNPPP 209
Qy 211 -HSNGFQSDYGNYPPOSQTLRSRSDDGYNWRKYQKQVKGSENPSPSYKCTYPCNCTKK 269
Db 210 VHYT-----QPSQYVREQKAEQYWRKYQKQVKGSENPSPSYKCTYPCNCTKK 260
Qy 270 VERSLDGQITEIVYKGTNHPKQONTNRNSNS--SSLAIPHNSIRTEIPDQSYATHGSG 328
Db 261 VERNLDGHTIETIVYKGNHHPKQSTRSSSIOQLA--YSN--LDITNQNAFLDNA 315
Qy 329 QMDSAAATPENSISIGDDDFEQSSQKCKSGGDEYDEDEPAKWKTEGENEGMSAPGSR 388
Db 316 QRSDFAGTNSASFGDEDIDQSPVSKGED--DGNPEAKWKGDNEVEISS--ASRT 372
Qy 389 VREPRVVVQTTSDIDLDDGYRWKTYGQKQVKGSENPSPSYKCTYPCNCTKKKVERASHD 448
Db 373 VREPRVVVQTTSDIDLDDGYRWKTYGQKQVKGSENPSPSYKCTYPCNCTKKKVERASHD 432
Qy 449 LRAVITYYEGKKNHVPAAARGSGSHSVNRPMPNPNASNHTNTAATSVRLLPVTHQ 504
Db 433 LRAVITYYEGKKNHVPAAARGSGSYAMNKPSPSSNS-----MPVPRPMLAN 482
Qy 505 NSLQ-----NQRSAAPPEQSGFTLEMLQSPGSGFGNPMQSYUNQOQLSDNV 555
Db 483 NSMQGNFNDTFNTFVQT--TQNPPTITLQMLQSSGSSSYSGPDTSSGSMNQMSNI 541
Qy 556 FSSRTKEPRDDMF 569
Db 542 -KPIITKEPKDDPF 554

RESULT 3
O81639 PRELIMINARY; PRT; 515 AA.
ID O81639
AC O81639;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Zinc finger protein.
GN ZFP1.
OS Pimpinella brachycarpa.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; campanulids; Apiales; Apiaceae; Pimpinella.
OX NCBI_TaxID=45043;
RN [1]
RP SEQUENCE FROM N.A.
RA Cho J.-I., Lee K.-W.;
RT "Pimpinella brachycarpa zinc finger protein PbZFP1 (WRKY1) mRNA.";
RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF080595; AAC31956.1; -.
DR InterPro; IPR003657; WRKY.
DR Pfam; PF03106; WRKY; 2.
DR PROSITE; PS50811; WRKY; 2.
SQ SEQUENCE 515 AA; 57553 MW; 8BF05537C621B46A CRC64;

```

```

Query Match 45.2%; Score 1379.5; DB 10; Length 515;
Best Local Similarity 55.1%; Pred. No. 5.7e-89;
Matches 313; Conservative 57; Mismatches 123; Indels 75; Gaps 24;

Qy 28 MTTSFSDLLASPLDNNKPPQGGISERTGSGVPKFKSTPPPSLSPSPSPSYFAIPPG 87
Db 1 MTSSGLDLAQAQPNDDFGSNWGFENQ-----KTKSFANPQSPFPSPSPSYFS----- 50

```



```

QY 88 LSPAEILLDPSVLLNSNLLPPTTGAFVAQSFNWKSSGCGNQOIVKEEDKSFNSFQTR 147
Db 51 -----FLDSPIQNNYTVSSSGNGTMAQSF-----KEENQNFDSFPAQ 92
QY 148 SGPASSTATYQSSNVT-----QOPWSQEQATKQDNFSSGKGMKTESNSM 197
Db 93 S-RPASSTSS--SSFPANTNSVEESLKRKQCGWNFEEAKNNEFORFSPFM-TMNQANM 148
QY 198 QSPSPASVQTN-HSG--FOSDYGNYPPQSOTL-----SRRSDDGYNWRKYGQKQVKG 249
Db 149 --LSPMTWQANMQNSAAVLQNSINYAQSSQSSQTNRDQSKLDDGYNRKYGQKQVKG 206
QY 250 SENPRSYKCTYPCNCTKKKVERSLDQITEIVYKGNHHPKQNTNRNSNSSSLAIPH 309
Db 207 SENPRSYKCTYPCNCTKKKVERSLDQITEIVYKGNHHPKQNTNRNSNSSSLAIPH 309
QY 310 SNSIRTEIPQSVATHSGQMSAATPENSISIGDD--FEQSSQCKSGGDEYDEDEPD 368
Db 261 QNSIPT-MPETSILLE--NGHLEPVTTTPENSLSFGEDDLFEQSGMN-KQGGD--DENEPD 314
QY 369 AKRWKIEG-NEGMSAPGSRVREPRVVQTTSDIDLDGGRWKYGQKVGKGNPR 427
Db 315 AKRWKEYENNETMSSLSGRTVREPRVVQTTSDIDLDGGRWKYGQKVGKGNPR 374
QY 428 YKCTHPGCPVRKHVERASHDLRAVITYEGKHNHDPVPAARGSGSVNRPMPNNAHNT 487
Db 375 YKCTQVGCPCVRKHVERASHDLRAVITYEGKHNHDPVPAARGSGSVNRPMPNNAHNT 431
QY 488 NTAATSVRLLPVTHQSDNSLQNSORQAPEGQSPFTLEMLQSPGSGFGSGFGNPMQSY-V 546
Db 432 TSAPTAIR--PTTNYL-NPLQNRAQ--PANGQAPFTLEMLQSPGSGFGSGFGNPMQSY 487
QY 547 NQOQLSDNVFSSRTKEPRDDMELESIL 574
Db 488 NQOQASGQFST-AKDEPDVDSFFDSFL 514

```

RESULT 4

```

Q9SQ04 ID Q9SQ04 PRELIMINARY; PRT; 514 AA.
AC Q9SQ04;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE Zinc-finger type transcription factor WRKY1.
GN WRKY1.
OS Petroselinum crispum (Parsley) (Petroselinum hortense).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; campanulids; Apiales; Apiaceae; Petroselinum.
OX NCBI_TaxID=4043;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99400614; PubMed=10469648;
RA Eulgem T., Rushton P.J., Schmelzer E., Hahlbrock K., Somssich I.E.;
RT "Early nuclear events in plant defence signalling: rapid gene
RT activation by WRKY transcription factors.";
RL EMBO J. 18:4689-4699(1999).
DR EMBL; AF121353; AAD55974.1; -
DR InterPro; IPR001360; Glyco_hydro_1.
DR InterPro; IPR003657; WRKY.
DR Pfam; PF03106; WRKY.
DR PROSITE; PS00572; GLYCOSYL_HYDROL_F1_1; 1.
DR PROSITE; PS00811; WRKY.
SQ SEQUENCE 514 AA; 57321 MW; 279064A5FF85D8D5 CRC64;

```

Query Match 44.0%; Score 1342; DB 10; Length 514;

Best Local Similarity 52.2%; Pred. No. 2.5e-86;

Matches 236; Conservative 74; Mismatches 123; Indels 74; Gaps 22;

QY 28 MTTFSDDLASPLDNNKPPQGLSERTGSGVFKSTPPPLPLSPSPSSYFAIPGP 87

Db 1 MSSSLGDLAQPNNDDFGSNWGFENQ-----NFKSFANGQLFPSPVSPSSYFS---- 50

```

QY 88 LSPAEILLDPSVLLNSNLLPPTTGAFVAQSFNWKSSGCGNQOIVKEEDKSFNSFQTR 147
Db 51 -----FLDSPIQNNYTVSSSGNGTMAQSF-----KEENQNFDSFPAQ 92
QY 148 SGPASSTATYQSSNVT-----QOPWSQEQATKQDNFSSGKGMKTESNSM 201
Db 93 S-RPASSTSS--SSFPANTNSVEESLKRKQCGWNFEEAKNNEFORFSPFM-TSDIATLQRTS 151
QY 202 PEIASVQTNHNS--GFQSDYGNYPPQSOTL-----SRRSDDGYNWRKYGQKQVKGSE 251
Db 152 PEWT--MNHANMQNSAAVLQNSINYAQSSQSSQTNRDQSKLDDGYNRKYGQKQVKGSE 208
QY 252 SENPRSYKCTYPCNCTKKKVERSLDQITEIVYKGNHHPKQNTNRNSNSSSLAIPH 311
Db 209 SENPRSYKCTYPCNCTKKKVERSLDQITEIVYKGNHHPKQNTNRNSNSSSLAIPH 262
QY 312 SIRTETIPQSVATHSGQMSAATPENSISIGDD--FEQSSQCKSGGDEYDEDEPD 370
Db 263 SIGT-MPETSILLE--NGRSEPVTTTPENSLSFGEDDLFEQSGMN-KPGDD--DENEPSDK 316
QY 371 RKWIEG-NEGMSAPGSRVREPRVVQTTSDIDLDGGRWKYGQKVGKGNPRSY 429
Db 317 RKWGYESNEPMSLSGRTVREPRVVQTTSDIDLDGGRWKYGQKVGKGNPRSY 376
QY 430 KCTHPGCPVRKHVERASHDLRAVITYEGKHNHDPVPAARGSGSVNRPMPNNAHNT 488
Db 377 KCTQVGCPCVRKHVERASHDLRAVITYEGKHNHDPVPAARGSGSVNRPMPNNAHNT 431
QY 489 NTAATSVRLLPVTHQSDNSLQNSORQAPEGQSPFTLEMLQSPGSGFGSGFGNPMQSY-VN 547
Db 432 TSAPTAIR--PTTNYL-NPLQNRAQ--PANGQAPFTLEMLQSPGSGFGSGFGNPMQSY 487
QY 548 NQOQLSDNVFSSRTKEPRDDMELESIL 574
Db 488 NQOQASGQFST-AKDEPDVDSFFDSFL 513

```

RESULT 5

```

Q40827 ID Q40827 PRELIMINARY; PRT; 514 AA.
AC Q40827;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE WRKY1.
OS Petroselinum crispum (Parsley) (Petroselinum hortense).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; campanulids; Apiales; Apiaceae; Petroselinum.
OX NCBI_TaxID=4043;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97051827; PubMed=8896462;
RA Rushton P.J., Torres J.T., Parniske M., Wernert P., Hahlbrock K.,
RA Somssich I.E.;
RT "Interaction of elicitor-induced DNA-binding proteins with elicitor
RT response elements in the promoters of parsley PrL genes.";
RL EMBO J. 15:5690-5700(1996).
DR EMBL; U48831; AAC49527.1; -
DR TRANSFAC; T03718; -
DR InterPro; IPR001360; Glyco_hydro_1.
DR InterPro; IPR003657; WRKY.
DR Pfam; PF03106; WRKY.
DR PROSITE; PS00572; GLYCOSYL_HYDROL_F1_1; 1.
DR PROSITE; PS00811; WRKY.
SQ SEQUENCE 514 AA; 57206 MW; 3B03C8A7DA486638 CRC64;

```

Query Match 43.6%; Score 1331; DB 10; Length 514;

Best Local Similarity 52.5%; Pred. No. 1.5e-85;

Matches 296; Conservative 72; Mismatches 128; Indels 68; Gaps 22;

QY 28 MTTFSDDLASPLDNNKPPQGLSERTGSGVFKSTPPPLPLSPSPSSYFAIPGP 87

```

Db 1 MSSLGDLAQPNDDFGSNWGFNQ-----KTKSFANQSLPSPFPVSPSSYFS----- 50
Qy 88 LSPAEILDSPVLLNSNLSPTTGAFAVQSFNWKSSGGNQOIVKEEDKSFNSFQTR 147
Db 51 -----FLDSPIONNYTIVSSSGNGTMAQSF-----KEENQNFDSFPPAQ 92
Qy 148 GPPASSTATYQSSNVTV-----QTQOPWFSQBATKQDNFSSGKGMKTSNSSMSQFS 201
Db 93 TRPASSTSSSLIPANTNLVDSLRKKGGMWFEPAPKNDPLMDNASV-TSDIATLQRI 151
Qy 202 PEIASVQTN-HSN-GFQSDYGNYPQSQTL-----SRRSDGYNWKYQKQVKGSENPR 254
Db 152 PEMTMQANMQSNAALQSNLNNYAQSQSQTNRDQSKLDDGYNWKYQKQVKGSENPR 211
Qy 255 SYIKCTVPNCTKKKVERSLDQITEIVYKGTTHNHPKQNTRRNSSSSSLAIPHSNIS 314
Db 212 SYIKCTYLNCTKKKVVETTFDGHTEIVYKGNHNHPKQNTRRSSQS-----YQNSIG 265
Qy 315 TEIPDQSVATHGSGQMSAATPENSISIGDD--FEQSOKCKSGGDEYDEDPDAKRWK 373
Db 266 T-MPESSULE--NKRSPVTTTENSISLFGEDDLFEQSMY-KPGDD--DGNEDPSRWK 319
Qy 374 IEGE-NEGMSAPGSRVREPRVVQTTSDIDILDDGYRWKYGKQVKGPNPRSYIKCT 432
Db 320 GEYESNEPMSLSGSRVREPRIVVQTTSDIDILDDGYRWKYGKQVKGPNPRSYIKCT 379
Qy 433 HPGCPVKRKHVERASHDLRAVITTYEGKHNHDVPAARGSGH-SVNRPMNNAASHHTAA 491
Db 380 QVGCPCVKRKHVERASHDLRAVITTYEGKHNHDVPAARGSGSYPAVNRP-----SDNTTAP 434
Qy 492 TSVRLLPVTHQSDNSLQNSORQAPEGSPTTLEMLQSGSFGSGFNPMQSY-VNQOQ 550
Db 435 TAIR--PTTNYL-NPLONPRAQ-PANGQAPPTLEMLQRPSEYFGTNTNTNTAINQ 490
Qy 551 LSDNVFSRTEKEPRDDMFLESLL 574
Db 491 QASGQFST-AKDEPDVDSFDSFL 513

```

RESULT 6

```

Q9XJ26 ID Q9XJ26 PRELIMINARY; PRT; 477 AA.
AC O9XJ26;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE NtWKY1.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamiales; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RS SEQUENCE FROM N.A.
RC STRAIN=cv. Xanthi;
RA Shinghi H., Yamamoto S., Suzuki K.;
RT "Analyses of an elicitor-responsive element and transcription factors
in cultured tobacco cells."
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB022693; BAA82107.1; -.
DR InterPro; IPR003657; WRKY.
DR Pfam; PF03106; WRKY. 2.
DR PROSITE; PS0811; WRKY. 2.
SQ SEQUENCE 477 AA; 53053 MW; 24B8617DBCS9B7A9 CRC64;

```

```

Query Match 41.9%; Score 1278; DB 10; Length 477;
Best Local Similarity 53.9%; Pred. No. 7.1e-82;
Matches 279; Conservative 64; Mismatches 109; Indels 66; Gaps 17;
Qy 71 LSPPPIPSSSYFAIPGLSPAELDSDPVLNNSNLSPTTGAFAVQSFNWKSSGGNQ 130
Db 2 ISSSPAPSSSYLAIPPLSPVLLDSPVLNNSNLSPTTGSF-----GNLN 49

```

```

Qy 131 IVKEEDKSFNSFQTRSGPPASSTATYQSSN-----VTVQTPWFSQBATKQDNFSSG 185
Db 50 -SKEDNSRISDFSFQSKA---ATSSSMFQSSAPRNSLEDLMTQK---QHANQNEPSTV 101
Qy 186 KGMKMTENSSMSQSPFPEIASVQTNHNSGFSQSDYGNYPQSQTLRRSDGYNWKYQK 245
Db 102 KTRGVKSEVAPIOQSFQSE-----KMQSNPAPVHYTH--PSQYVREQKAEDGYNRKYQK 154
Qy 246 QVKGSENPRSYIKCTVPNCTKKKVERSLDQITEIVYKGTTHNHPKQNTRRNSSNS-SS 304
Db 155 QVKGSENPRSYIKCTFPNCTKKKVERNLQGHTEIVYKGNHNHPKQNTRRSSSQIQN 214
Qy 305 LAIPHNSIRTEIPDQSVATHGSGQMSAATPENSISIGDDDFEQSSQCKSGGDEYDE 364
Db 215 LA--YSN--LDITNQSNFLDNAQRDSFAGTDSNSSFQDEIDQGSPIKSGED--DG 267
Qy 365 DEPDAKRWKIEGNEGMSAPGSRVREPRVVQTTSDIDILDDGYRWKYGKQVKGPN 424
Db 268 NEPEPKRWKGDNEVEISS-ASRTVREPRIVVQTTSDIDILDDGYRWKYGKQVKGPN 326
Qy 425 PRSYIKCTHPCPVKRVKVERASHDLRAVITTYEGKHNHDVPAARGSGSHSVNRPMNNA 484
Db 327 PRSYIKCTFTGCPVKRKHVERASHDLRAVITTYEGKHNHDVPAARGSGSYAMNRP 386
Qy 485 NHTNTAATSVRLLPVIHQ-----SDNSLQ-----NORSQAPPEGQSPFTLEMLQSPG 531
Db 387 NS-----MPVPRPSMLANNNOGMNFNTFTFTRVQT-TQNQPPITLQMLQSSG 435
Qy 532 SFQSGFGNPMQSYVNVQQLSDNVFSRTEKEPRDDMF 569
Db 436 NSSYSGFDTSYGSGSYNMQMNMNI-KPISKEEPKDDFF 472

```

RESULT 7

```

Q9SXP5 ID Q9SXP5 PRELIMINARY; PRT; 353 AA.
AC Q9SXP5;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Transcription factor NtWKY2.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamiales; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RS SEQUENCE FROM N.A.
RC STRAIN=cv. Xanthi;
RA Shinghi H., Yamamoto S., Suzuki K.;
RT "Analyses of an elicitor-responsive element and transcription factors
in cultured tobacco cells."
RL Submitted (NOV-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB020590; BAA7383.1; -.
DR InterPro; IPR003657; WRKY.
DR Pfam; PF03106; WRKY. 2.
DR PROSITE; PS0811; WRKY. 2.
SQ SEQUENCE 353 AA; 40061 MW; 33E6D3A1FF21317E CRC64;

```

```

Query Match 40.1%; Score 1222; DB 10; Length 353;
Best Local Similarity 65.3%; Pred. No. 4.1e-78;
Matches 245; Conservative 42; Mismatches 60; Indels 28; Gaps 10;

```

```

Qy 207 VQTNHNSGPOSYGVNTPPQSQTLRRSDGYNWKYQKQVKGSENPRSYIKCTYPCNCT 266
Db 1 MQSLKXNGQSNQYNNQSSQSIREQKRSDDGYNWKYQKQVKGSENPRSYIKCTFPNCT 60
Qy 267 KKXVERSLDQITEIVYKGTTHNHPKQNTRRNSSSSSLAIP---HNSNIRTEIPDQSYA 323
Db 61 KKXVERSLDQITEIVYKGNHNHPKQNTRRSSSTASSLTPLOYTKLMKSQI-----IS 116
Qy 324 THSGQWMSAATPENSISIGDD--DFEQSSQCKSGGDEYDEDPDAKRWKIEGENEGM 381

```

```

Db 117 SYGNS-----ATPENSISFGDDHDEQSSRSRSGDDNEEBEEDPSKRWKRESEGL 171
QY 382 SAGSRTVRPRVVVQTTSIDILDDGYRWKYGKQVKGNNPNPRSYKCTHGPCPVKX 441
Db 172 SAGSRTVRPRVVVQTTSIDILDDGYRWKYGKQVKGNNPNPRGYKCTSGCPVKX 231
QY 442 VERASHDLRAVITYEGKHNHVPAAAGSGSHSVNRPMPNASHNTAATSVRLPLVIH 501
Db 232 VERASQDIRSVITYEGKHNHVPAAAGSG---INRPVAFNIT--YNNGANAMAIRPSV- 285
QY 502 QSDNSLQNSQAPPEGOSPTLEMLQSPGF-GFSGFGNPMQSYVNOQLSDNVSSRT 560
Db 286 TSQIPLQSIKRPQ-----QSPFTLEMLKHPNSYNGFGSYVNSEDSYEN--QLQDNVFSRA 338
QY 561 KEPRDDMFLESLLC 575
Db 339 KDEPRDDMFMETLLC 353

RESULT 8
Q941B4 PRELIMINARY; PRT; 378 AA.
ID Q941B4
AC Q941B4; 2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE WRKY DNA-binding protein (Fragment).
GN NTRKY-8.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamids; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Petit Havana SRI;
RA Maeo K., Hayashi S., Kojima-Suzuki H., Morikami A., Nakamura K.;
RT "Role of conserved residues of the WRKY domain in the DNA-binding
RT activity of tobacco WRKY family proteins.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB063575; BAB61055.1; -
DR InterPro; IPR003657; WRKY.
DR Pfam; PF03106; WRKY; 2.
DR PROSITE; PS50811; WRKY; 2.
KW DNA-binding.
FT NON TER.
SQ SEQUENCE 378 AA; 43050 MW; DA0BC909C078413D CRC64;

Query Match 39.4%; Score 1201.5; DB 10; Length 378;
Best Local Similarity 56.8%; Pred. No. 1.3e-76;
Matches 254; Conservative 48; Mismatches 68; Indels 77; Gaps 15;

```

```

QY 137 KPSFNFSQTRSGPPASSTAYTSSNVVTOQPSWFOEATKQDNFSGGKGMKNTSSS 196
Db 1 KNYSDFSFQ---PQFASVPSQTNVPVL-GKQSNYQESRKQND-----ENAG 45
QY 197 MQSFSPPIASVOTNHSNGFSDYGNYPQSQTLSSRRSDCGYNWRKYQKQVKGSENPRSY 256
Db 46 -----TSLEQLKNGNQSNYKQ-----SSRSEDGYNWRKYQKQVKGSENPRSY 91
QY 257 YKTYPCNPTKKKVERSLDQGIETIVYKGTNHPKP-QNTRRNSSNSSSLAIPHSNIRT 315
Db 92 YKTYPCNPTKKKVERSLDQGIETIVYKGTNHPKPQTQSTR-----SSSLAIPYNTQTN 147
QY 316 EIPDQSVATHSGQMDAAATPENSISIGDDDFEQSKCKSGGDEYDED-----EPDA 369
Db 148 EIPDQV-----STPENSISIFGDDDEKS-----RSRGDDDFDEEEDPSKEPDP 191
QY 370 KRWKIEGECMSAPGSTRVREPRVVVQTTSIDILDDGYRWKYGKQVKGNNPNPRSY 429
Db 192 KRWKRESEGLSAPGSTRVREPRVVVQTTSIDILDDGYRWKYGKQVKGNNPNPRSY 251

```

```

QY 430 KCTHPCPVKRVKVERASHDLRAVITYEGKHNHVPAAAGSGSHSVNRPMPNASHNT 489
Db 252 KCTSPGCPVVKVERASQDIRSVITYEGKHNHVPAAAGSG---AINRPVA-----PTIT 303
QY 490 KATSVELLPVIHQSDNSLQNSQAPPEGOSPTLEMLQSPGF-GFSGFGNPMQSYVNO 548
Db 304 YNNAIPRPSV-----TSQIPLQSQPTLEMLKHPNSYNGFGSYATSEDSYEN- 352
QY 549 QQLSDNVFSRRTKEPRDDMFLESLLC 575
Db 353 -QLQDNVFSRAKDEPRDDMFMETLLC 378

RESULT 9
Q888P5 PRELIMINARY; PRT; 512 AA.
ID Q888P5
AC Q888P5;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE Putative WRKY-type DNA binding protein.
GN AT2G38470.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Rounsley S.D., Lin X., Ketchum K.A., Crosby M.L., Brandon R.C.,
RA Sykes S.M., Kaul S., Mason T.M., Kerlavage A.R., Adams M.D.,
RA Somerville C.R., Venter J.C.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Lin X.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RA Town C.D., Kaul S.;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC004683; AAM14994.1; -
DR InterPro; IPR003657; WRKY.
DR Pfam; PF03106; WRKY; 2.
DR PROSITE; PS50811; WRKY; 2.
SQ SEQUENCE 512 AA; 56457 MW; 8F19CBE41BC18662 CRC64;

Query Match 37.4%; Score 1139.5; DB 10; Length 512;
Best Local Similarity 44.7%; Pred. No. 4.4e-72;
Matches 254; Conservative 80; Mismatches 144; Indels 90; Gaps 15;

```

```

QY 40 LDNNKPPQ-----GGLSERTGS-----GYPKFKSTPPPLPLSPPISSSYPAIP 85
Db 1 MDNSRTQNMNGSANWSQQSGRTSTSSLEDLEIPKFRSPAPSSISPSLVSPSTCF- 57
QY 86 PGLSPAEILDSPVLNLS--NILPSPTTGAPVAOSFNWSSSSGNGNQIVKEEDKFSN- 141
Db 58 ---SPSLFLDSPAFVSSSANVLASPTTGALI-----TNVTNKGINEGDKSNNNFN 106
QY 142 ---FSQTSRGGPPASSTAYTSSNVVTOQPSWFOEATKQDNFSGGKGMKNTSSSMQ 198
Db 107 LDFSPHTQSSGVSAPTTTTTTTTTTTTTTNS--SIFSQEQQKKKQSEQWSTETRPNQ 164
QY 199 SFSPEIASVQTNHNSNGFSDYGNYPQSQTLSSRRSDCGYNWRKYQKQVKGSENPRSY 258
Db 165 AVS-----YNGRE-----QRKGEDGYNWRKYQKQVKGSENPRSY 201
QY 259 CTYVNCPTKKKVERSLDQGIETIVYKGTNHPKPQNTRRNSSNSSSL-----AIPHSN 311
Db 202 -CTFPNCPTKKKVERSLDQGIETIVYKGTNHPKPQNTRRNSSNSSSLFHSVAVNASLDHNR 261

```

QY 312 SIRTEIPDQSYATHGSGQMSAATPENSSTISIGDDDFEQSSQKCKSGGDEYDEPDQAKR 371
 Db 262 QASDQPNNSNFHQSDSGFQMQEDNTTSDSGDDEFEQGS-SIVSRDEDDCSEPEAKR 320
 QY 372 WKIEGENEGMSAPGSRVREPRVVQVQTSIDILDDGYRWKYGQKVKGNPNPRSYK 431
 Db 321 WKGNETNGGSGSKTVREPRVVQVQTSIDILDDGYRWKYGQKVKGNPNPRSYK 380
 QY 432 THPGCPVKHVERASHDLRAVITTYEGKHNDVPAARGSGSHSVNRPMPNNAHNTTAA 491
 Db 381 TTIGCPVKHVERASHDMRAVITTYEGKHNDVPAARGSG-YATNRAPODSSVPTRPAA 439
 QY 492 TSVRLLPVHQSDNSLQNSQAPPEQSPFTLEMLQ---SPSGSGSGFGNPMQSYVN 547
 Db 440 IA-----GHSNVTTSQAPYTLQMLHNNNTNTGPGFYAMNNNNNSLQ 483
 QY 548 QQQLSDNVFSRTKEPRDD-MFLESLL 574
 Db 484 TQNFVGGGFSRAKEEPEETSFFDSFM 511

RESULT 10

Q9XFF1 ID Q9XFF1 PRELIMINARY; PRT; 501 AA.
 AC Q9XFF1;
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE DNA-binding protein WRKY1.
 GN WRKY1.
 OS Avena sativa (Oat).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
 OC Aveneae; Avena.
 OC NCBI_TaxID=4498;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Rhiannon; TISSUE=aleurone;
 RA Screen S.E., Powell R., Rollason R., Huttly A.K., Lazarus C.M.;
 RT "Isolation of WRKY-type DNA-binding proteins from Avena sativa."
 RL Submitted (Apr-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF140554; AAD32677.1; -.
 DR InterPro: IPR003657; WRKY.
 DR Pfam: PF03106; WRKY; 2.
 DR PROSITE: PS50811; WRKY; 2.
 KW DNA-binding.
 SQ SEQUENCE 501 AA; 53816 MW; E9C74BE8ED21E4D8 CRC64;

Query Match 36.6%; Score 1116.5; DB 10; Length 501;
 Best Local Similarity 46.0%; Pred. No. 1.8e-70;
 Matches 256; Conservative 60; Mismatches 135; Indels 105; Gaps 19;
 QY 64 TPPPSLPLSPPISSPSYFAIPPGLSPAELLDSVLLNSNLPSPITGAFVAQSFNW-- 121
 Db 4 TPTTSP---PPASPSYF-----NNMSTGFLDSILLTPS-LFSPITGSPFPLEPLNWMG 54
 QY 122 ---KSSGGGNOQIVKE-EDKSFNSFQTRSGP-PASSTAT-----YQSS----- 161
 Db 55 TAPESNDGLQLGSKVDGQOQROYSQGTFTTAAAPVPAAMPNTTNTASSFLQSSMPMAQQGH 114
 QY 162 -NVTVQTPQKSPQEAATKQDNFFSGKGMKMTENSQSSQSPETASVQTNHSGFQSDYG 220
 Db 115 DSYTGEQQPMYSYQDAGSMD-----AMTTRPASFTPTYEAPDMVG-NGGYNNAAPVSSG 167
 QY 221 NYPQSQOTLSRR--SDDGYNWKYGQKVKGNPNPRSYKCTYPCNCTKKKVERSLDQGI 278
 Db 168 TTAGYGRVQSRPPSSDDGYNWKYGQKVKGNPNPRSYKCSFAGCSTKKKVEQAPDGV 227
 QY 279 TEIVYKGTNHHPKQNTNRSSSSSLAIPHNSIRTEIPDQSYATHGSGQMSAATPEN 338
 Db 228 TEIVYKGTNHHPKQNTNRSSAPASS-----SYASPDASSDALSGTPEN 271

QY 339 SSISIGDDDFEQSSQ---KCKSGGDEYDEPDQAKRWKIEGENEGMSAPGSRVREPRV 394
 Db 272 SSASYGDDDTNGVSSALACQFGGGEFADNEPDSKRWRTDSDAEGVPVGANRTVREPRV 331
 QY 395 VVQVTSIDILDDGYRWKYGQKVKGNPNPRSYKCTHFGCPVKHVERASHDLRAVIT 454
 Db 332 VVQVTSIDILDDGYRWKYGQKVKGNPNPRSYKCTTAGCPVKHVERASHDLRAVIT 391
 QY 455 TYEGKHNDVPAARGSGSHSVNRPMPNNAHNTTNTAATSVRLLPVTHQSDNSLQ----- 508
 Db 392 TYEGKHNDVPAALRGSA-----AAARYRAAPM--QAASYLQGGGGY 432
 QY 509 -----NQRSAPEQSGSPFTLEMLQSPGSGFGNPMQSYVNQOQLSDNVF--SS 558
 Db 433 SSLRPDGFGGGAQAPADQGFALS-----GPD-YNNSSYSYSGMQQNDAMYDAA 483
 QY 559 RTKEPRDDMFLESLL 574
 Db 484 RTKDEPRDDMFESQL 499

RESULT 11

Q94D89 ID Q94D89 PRELIMINARY; PRT; 357 AA.
 AC Q94D89;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
 DE Putative DNA-binding protein ABF1.
 GN P0439E11.17.
 OS Oryza sativa (Rice).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
 OC Ehrhartoidae; Oryzeae; Oryza.
 OC NCBI_TaxID=4530;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=cv. Nipponbare;
 RA Sasaki T., Matsumoto T., Yamamoto K.;
 RT "Oryza sativa nipponbare (GA3) genomic DNA, chromosome 1, PAC
 clone: P0439E11.";
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AP003315; BAB61266.1; -.
 DR Gramene: Q94D89; -.
 DR InterPro: IPR003657; WRKY.
 DR Pfam: PF03106; WRKY; 2.
 DR PROSITE: PS50811; WRKY; 2.
 KW DNA-binding.
 SQ SEQUENCE 357 AA; 38867 MW; E23FE44F5B550854 CRC64;

Query Match 35.4%; Score 1081; DB 10; Length 357;
 Best Local Similarity 59.4%; Pred. No. 3.5e-68;
 Matches 227; Conservative 28; Mismatches 73; Indels 54; Gaps 11;
 QY 216 QSDYGNVYPOSOTLSRRSDDGYNWKYGQKVKGNPNPRSYKCTYPCNCTKKKVERSLD 275
 Db 6 QCGGGYSQSQSQ--RRSDDGYNWKYGQKVKGNPNPRSYKCTYPCNCTKKKVERSLD 63
 QY 276 GQTEIVYKGTNHHPKQNTNRSSSSSLAIPHNSIRTEIPDQSYATHGSGQMSAAT 335
 Db 64 GQTEIVYKGTNHHPKQNTNRSSSSAAQVLSQSG-----DMSEHSFG-QMSGTAAT 115
 QY 336 PENSSISIGDDDFEQSSQK-CKSGGDEYDEPDQAKRWKIEGENEGMSAPGSRVREPRV 394
 Db 116 PENSSASFGDDEIRVSGSPRAGNGGGDEFFDDPSKRWKIDGGEGSLMAGNRTVREPRV 175
 QY 395 VVQVTSIDILDDGYRWKYGQKVKGNPNPRSYKCTHFGCPVKHVERASHDLRAVIT 454
 Db 176 VVQVTSIDILDDGYRWKYGQKVKGNPNPRSYKCTTAGCPVKHVERASHDLRAVIT 235
 QY 455 TYEGKHNDVPAARGSGSHSVNRPMPNNAHNTTNTAATSVRLLPVTHQSDNSLQ--RS 512
 Db 236 TYEGKHNDVPAARGSA--ALYRPAAPAA-----AATS-----SHFPLNPQPPM 278

```
Qy 513 QAPPEGSPFTLEMLQSPGSGF-----GFSGF-----GNPMQSYVNOQOOLS 552
Db 279 SYQPTGQPVYALR-----PDGFGGQCPFGVVGSGSFGSFGSFGDDARGSYMSQHQQOQRON 334
Qy 553 DNVFSSRTKEPRDDMFLESLL 574
Db 335 DAMHASRAKEBEPGDDMPFQNSL 356

RESULT 12
Q38704
ID Q38704 PRELIMINARY; PRT; 402 AA.
AC Q38704;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE DNA-binding protein (Fragment).
OS Avena fatua.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
OC Avenae; Avena.
OX NCBI_TaxID=4499;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FA; TISSUE=Aleurone;
RX MEDLINE=96128012; PubMed=8541496;
RA Rushton P.J., Macdonald H., Huttly A.K., Lazarus C.M., Hookey R.;
RT "Members of a new family of DNA-binding proteins bind to a conserved
RT cis-element in the promoters of alpha-amyl 2 genes.";
RL Plant Mol. Biol. 29:691-702(1995).
DR EMBL; Z48429; CAA88326.1; -
DR TRANSPAC; T03726; -
DR InterPro; IPR003657; WRKY.
DR Pfam; PF03106; WRKY; 2.
DR PROSITE; PS50811; WRKY; 2.
KW DNA-binding.
FT NON TER
SQ SEQUENCE 402 AA; 43232 MW; DA81C09530E9A868 CRC64;

Query Match 33.98; Score 1032.5; DB 10; Length 402;
Best Local Similarity 53.2; Pred. No. 1.1e-64;
Matches 225; Conservative 48; Mismatches 105; Indels 45; Gaps 15;

Qy 168 QPMFQEAATKQDNFSSGKGMKMTSSSMQSFPEIASVQTNHNSGFQDYGNYPPQSQ 227
Db 2 QQPMGYQQ-----PGAM--DAGANAASF--APAVQATSE--MAPSGGVYRQTH 46
Qy 228 TLRRSDGYNWRKYQKQKVGSENPRSYKCTYPNCPTKKKVERSLDGQITRIVYKGT 287
Db 47 SQRSSDDGYNWRKYQKQKVGSENPRSYKCTFPNCPTKKKVTSTEGQITRIVYKGT 106
Qy 288 NHPKPQNTRRNSSSSSLAIPHNSIRTEIPDQSYATHSGQMDSA--ATPENSISIGD 345
Db 107 NHAKPLSTRSGGGGGGGA-----QVLSQGGGGDASEHSGAMGAPVSTPENSASF 162
Qy 346 DDFEQSQCKKS--GGDYDEDPDAKRWKTEGENEGM--SAPGSRVTREPRVVVQTTSDID 403
Db 163 DETGASSPRAGNVGGDDLDDEPDSKWRKDGDEGSGNSVAGNRTREPRVVVQTTSDID 222
Qy 404 ILDDGYRWKYQKQKVGKGNPNRPSYKCTHFGCPVRKHVERASHDLRAVITTYEGKHNH 463
Db 223 ILDDGYRWKYQKQKVGKGNPNRPSYKCTTVGCPVRKHVERASHDLRAVITTYEGKHNH 282
Qy 464 VPAARGSGSHSVNRPMPNNSN-----HTNTAATSVRLLP--VIHQSDNSLQNRSQAPPEG 518
Db 283 VPAARGSA--ALYRPAPRAAADSAMSTSQQYTANQQQPSAMTYQTSAAGTQQYAPRPDG 340
Qy 519 -----QSPFTLEMLQSPGSGSFGS--GFGNPMQSYV-----QQQLSDNVFSSRTKEEPRDDM 568
Db 341 FGSONQSGSFGFN-----GSFGSAGFNDPTGTSYMSQHQQOQRONDAMQASGAKEEPRDM 395
Qy 569 FLE 571
```

```
Db 396 FFQ 398

RESULT 13
Q94IB3
ID Q94IB3 PRELIMINARY; PRT; 727 AA.
AC Q94IB3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE WRKY DNA-binding protein.
GN NWRKY-9.
OS Nicotiana tabacum (Common tobacco).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamids; Solanales; Solanaceae; Nicotiana.
OX NCBI_TaxID=4097;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Petit Havana SRI;
RA Maao K., Hayashi S., Kojima-Suzuki H., Morikami A., Nakamura K.;
RT "Role of conserved residues of the WRKY domain in the DNA-binding
RT activity of tobacco WRKY family proteins.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB063576; BAB61056.1; -
DR InterPro; IPR003657; WRKY.
DR Pfam; PF03106; WRKY; 2.
DR PROSITE; PS50811; WRKY; 2.
KW DNA-binding.
SQ SEQUENCE 727 AA; 78975 MW; A5674A4F282F6F24 CRC64;

Query Match 30.4%; Score 927; DB 10; Length 727;
Best Local Similarity 38.1%; Pred. No. 6.3e-57;
Matches 237; Conservative 79; Mismatches 154; Indels 152; Gaps 23;

Qy 41 DNKKPPQGGGLSE-----RTGSGVPEKF--KSTPPSLPLSPPIPSPIFYAIPGLSPAELL 94
Db 74 DQKSSRGGLERMAARAGFNAPKLNTESLRPADLSQNGVRSP--YLTIPPLGLSPTSL 131
Qy 95 DSVILNSNLLPSPTTGAFVAQSFNWKSSGGNQIQIVKEEDK-----SFSNPSF- 144
Db 132 ESPVFLSNLQVSPPTGKFP--QFASGIESRNSTFMEDPDPRKKNALLESINSSSFSFK 188
Qy 145 ---OTRSGPPASSTATYQGSNVT-----VOTQCPWSFOE 175
Db 189 PVDETAPSLFPGTTSRVNSNTISQCFPIKIVSVHSQNSLLSHSVATQMTQSEKGLHQ 248
Qy 176 ATKQDNFSSGKGMKMTSSSMQSFPEIASVQTNHNSGFQSDYGNYPPOSQTLRR--- 232
Db 249 SSDPFRFSAEKGVDRDSNVTPESRNFQ---SVGSNMEHSGPPLD---EPQDEIDQKRVGG 300
Qy 233 -----SDDGYNWRKYQKQKVGSENPRSYKCTYPNCPTKKKVERSLDGQITRIVYK 284
Db 301 DPNVVGAPAEADGYNWRKYQKQKVGSEYPRSYKCTHPNCPVKKKVERSHGHTIYIK 360
Qy 285 GTNHHPKPQNTRRN---SNS-----SSLAIPHNSNI--RTEI---PD----- 319
Db 361 GANHHPKPPNRRASALGSTNSLGLQLDGAEQVSGSNGDLGRANTQKAPDAGGLDWRNN 420
Qy 320 -----QSYATHSGQMDS--AATPENSISISIGDDDFEQSQCKSGGD 360
Db 421 NLDVTSSAHLGSAYCNCSASFVQNNVTQLESQGVADVSTFSNDEDDDRGTHGSVSOY 480
Qy 361 EYDEDEPDAKRWKIEGENEGMSAPGSRVTREPRVVVQTTSDIDILDDGYRWKYQKGVK 420
Db 481 DGEDESESKRRKLETYSTDMSC--ATRAIREPRVVVQTTSEVDILDDGYRWKYQKGVK 539
Qy 421 GNPNPSYKCTHFGCPVRKHVERASHDLRAVITTYEGKHNHDPVPAARGSGSHSVNRPMP 480
Db 540 GNPNPSYKCTAGCNVRKHVERASHDLKSLVITTYEGKHNHDPVPAARN----- 588
Qy 481 NNASNHTNTAATSVRLLP-----IHQSDNS--LQNRSQAPPEGSPFTLEMLQSPG 531
```

```
Db 589 ---SSHVNSASNTLPAPVTAPPAQSHLHPEPAQLQANAR-----FDRQPSLG 635
QY 532 SFGSGFGNPMQSY-VNQOOLS 552
Db 636 SFG-----PTGYSYGINQOGLA 653

RESULT 14
QYARC7
ID QYARC7 PRELIMINARY; PRT; 739 AA.
AC QYARC7;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE Hypothetical 79.9 kDa protein.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamids; Solanales; Solanaceae; Solanum.
OC NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21178822; PubMed=11283350;
RA Rosberg M., Theres K., Acarkan A., Herrero R., Schmitt T.,
RA Schumacher K., Schmitt G., Schmidt R.;
RT "Comparative sequence analysis reveals extensive microcolinearity in
RT the lateral suppressor regions of the tomato, Arabidopsis and Capsella
RT genomes.";
RL Plant Cell 13:979-988(2001).
DR EMBL; AJ303345; CAC36402.1; -.
DR InterPro; IPR003657; WRKY.
DR Pfam; PF03106; WRKY; 2.
DR PROSITE; PS50811; WRKY; 2.
KW Hypothetical protein.
SQ SEQUENCE 739 AA; 79854 MW; P9CAF86DA15B8722 CRC64;

Query Match 30.3%; Score 924; DB 10; Length 739;
Best Local Similarity 35.6%; Pred. No. 1.1e-56;
Matches 242; Conservative 95; Mismatches 177; Indels 166; Gaps 25;

QY 36 LASPLDNNKPPQGG-----LSERTGSGVPRF--KSTPPPSLPLSPSPSSYFAIPPL 88
Db 78 LSPSPDQKNSRGGLLIERIAARAGFNAPKLNTESLRPAQLRQNGVRSP--YLTIPPL 135
QY 89 SPAELDPSVLLNSNLPSTTGAFVAQS--FNWKSXSGGNGQIVKEEDKSF-----SNF 142
Db 136 SPTLLSPVFLNSLVQSPSTTGKFLFSSGIDNRNSALMMDTDKKNDALESINSSSF 195
QY 143 SFQ-----TRS-----GPPASSTATYQSSNVTVQTO-----Q 169
Db 196 SFKVPETAPSLPFGTTTSSRLQVNPSPNFQSGFPNIEVSHVNSQLVSHRVEATQNPQT 255
QY 170 PWSFOEATKQDNFSSGKGMKNTSSSQSFSPASVQTNHNGFQSDYGNYPQSQTL 229
Db 256 NGTLQSSDPRFSAEKDVM--ANNVLDSTFTQVGSADVHSPPLDE-----PQEDI 307
QY 230 SRR-----SDDGYNRWYKQKQVKGSENPSPYKCTYPCNCTPKKKKVERSLDQI 278
Db 308 DQGGGDPNAGAPAEADGYNRWYKQKQVKGSENPSPYKCTHTPTCPVKKKVERSPGHI 367
QY 279 TEIVYKGNHHPKQNTK-----
Db 368 TEIYKGAHNHPKPPNRRSALGSTNSLGLDQVGAEGASGVNGDLQANFHKAPGGG 427
QY 297 ----RNSNSSLAIHNSNIRTEIPQSYATHSGQWDSA--ATPNSISIGDDDFEQS 351
Db 428 GFWNRNRRNLANLSEHNR-----SAPFSAQNTRLESQADVVSNSFNENDEDEDRG 481
QY 352 SQKCKSGGDEDEPDAKRWKIEGENEGMAGPSRTVREPRVVVQTTSDIILDGVRW 411
Db 482 THGVSQGYEGEGESESRRKLETYSADMTG--ATRAIREPRVVVQTTSEVTDIILDGVRW 540
```

```
QY 412 RYKQKVVKGNPNRSPRYKCTHPCPVRRKHVERASHDLRAVITTYEGKNHNDVPAARGSG 471
Db 541 RYKQKVVKGNPNRSPRYKCTHPCPVRRKHVERASHDLKSVITTYEGKNHNDVPAARNS- 599
QY 472 SHSVNRMPNPNASNTHTAATSVRLLPLVTHQSDNS--LQNR-----SQAPP 516
Db 600 SH-----VNSGASNTHPASVTAPAQNHILHPEAQLQANAMARFDRQPSLGSFGLSGRPP 653
QY 517 E-QQSP-FTLEMLQSPG--SFGSPF-----GN-PMQSYVNOQQLSDNVFSSRTKEEP 564
Db 654 QLGPNPFGSYGMNQGLSLAMAGFHPNQNKPEVPMHPYLGQPRMMDMGFMFPKEEP 713
QY 565 RQD-----MFLES 573
Db 714 KVEPLSDPGPTAYQQFLNSL 733

RESULT 15
QYARC7
ID QYARC7 PRELIMINARY; PRT; 739 AA.
AC QYARC7;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-OCT-2002 (Tremblrel. 22, Last annotation update)
DE Hypothetical 79.8 kDa protein.
OS Lycopersicon esculentum (Tomato).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
OC Asteridae; lamids; Solanales; Solanaceae; Solanum.
OC NCBI_TaxID=4081;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21178822; PubMed=11283350;
RA Rosberg M., Theres K., Acarkan A., Herrero R., Schmitt T.,
RA Schumacher K., Schmitt G., Schmidt R.;
RT "Comparative sequence analysis reveals extensive microcolinearity in
RT the lateral suppressor regions of the tomato, Arabidopsis and Capsella
RT genomes.";
RL Plant Cell 13:979-988(2001).
DR EMBL; AJ303343; CAC36397.1; -.
DR InterPro; IPR003657; WRKY.
DR Pfam; PF03106; WRKY; 2.
DR PROSITE; PS50811; WRKY; 2.
KW Hypothetical protein.
SQ SEQUENCE 739 AA; 79809 MW; 6D97C8E64C1F221B CRC64;

Query Match 30.2%; Score 922; DB 10; Length 739;
Best Local Similarity 36.0%; Pred. No. 1.5e-56;
Matches 243; Conservative 96; Mismatches 182; Indels 154; Gaps 25;

QY 36 LASPLDNNKPPQGG-----LSERTGSGVPRF--KSTPPPSLPLSPSPSSYFAIPPL 88
Db 78 LSPSPDQKNSRGGLLIERIAARAGFNAPKLNTESLRPAQLRQNGVRSP--YLTIPPL 135
QY 89 SPAELDPSVLLNSNLPSTTGAFVAQS--FNWKSXSGGNGQIVKEEDKSF-----SNF 142
Db 136 SPTLLSPVFLNSLVQSPSTTGKFLFSSGIDNRNSALMMDTDKKNDALESINSSSF 195
QY 143 SFQ-----TRS-----GPPASSTATYQSSNVTVQTO-----Q 169
Db 196 SFKVPETAPSLPFGTTTSSRLQVNPSPNFQSGFPNIEVSHVNSQLVSHRVEATQNPQT 255
QY 170 PWSFOEATKQDNFSSGKGMKNTSSSQSFSPASVQTNHNGFQSDYGNYPQSQTL 229
Db 256 NGTLQSSDPRFSAEKDVM--ANNVLDSTFTQVGSADVHSPPLDE-----PQEDI 307
QY 230 SRR-----SDDGYNRWYKQKQVKGSENPSPYKCTYPCNCTPKKKKVERSLDQI 278
Db 308 DQGGGDPNAGAPAEADGYNRWYKQKQVKGSENPSPYKCTHTPTCPVKKKVERSPGHI 367
QY 279 TEIVYKGNHHPKQNTK-----AIPHNSNIRTEIPDOSY--ATHGSG 328
Db 368 TEIYKGAHNHPKPPNRRSALGSTNSLGLDQVGAEGASGVNGDLQANFHKAPGGG 427
```

```
QY 329 QMD-----SAATPENSSISIGDDDFEQSSQCKS 357
Db 428 GFDWRNNLDANLGSEHCNRSAPFSAQNNTLRLESGDAVDVSSNFSNDEDEDDRGTHGSVS 487
QY 358 GGDEYDEDEDAKRWKIEGENEGNSAPGSRVPRVVQTTSDIDILDDGYRWRKYGOK 417
Db 488 QGYEGEGDESESKRRKLETYSADMTG--ATRAIREPRVVQTTSEVDILDDGYRWRKYGOK 546
QY 418 VVKGNPNPRGYKCTHPGCPVRKHVERASHDLRAVITTYEGKHNHDVPAARGSGSHSVNR 477
Db 547 VVKGNPNPRGYKCTSHAGCNVRKHVERASHDLKSATITTYEGKHNHDVPAARNS-SH---- 601
QY 478 PMPNNASNHTNTAATSVRLLPVIHQSDNS-LQNR-----SQAPPE-GQSP 521
Db 602 --VNSGASNTHPASVTAPAQNHLHRPEAAQLQANAMARFDRQPSLGSLGSRPPQLGPNP 659
QY 522 -FTLEMLQSPG--SFGSGF-----GN-PMQSYVNOQOLSDNVFSSRTKEEPRDD--- 567
Db 660 GFSYGMNQOQGLSLAMAGFHPNQNKPEVPMHPYLGQPRPMHDMGMFMFPKEPKVEPLS 719
QY 568 -----MELESLL 574
Db 720 DPGPTAYQQFNSLL 734
```

Search completed: January 20, 2004, 16:17:51
Job time : 44 secs

THIS PAGE BLANK (USPTO)